

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:20 ; Search time 51.7 Seconds
(without alignments)
58.008 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVMYTDGKWNDRNCLOSRLATCEF 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155	100.0	27	20	AAV25519
2	155	100.0	27	21	AAV77986
3	155	100.0	27	22	AAV63349
4	137	88.4	117	7	AAV60439
5	137	88.4	248	7	AAV60665
6	137	88.4	248	7	AAV60666
7	137	88.4	248	7	AAV60441
8	137	88.4	248	7	AAV60442
9	137	88.4	248	8	AAV70662
10	137	88.4	248	8	AAV70663
11	137	88.4	248	9	AAV80694

12	137	88.4	248	9	AAV82980
13	137	88.4	248	11	AAV05092
14	137	88.4	248	11	AAV05091
15	137	88.4	248	11	AAV04215
16	137	88.4	248	11	AAV04216
17	137	88.4	248	11	AAV06331
18	137	88.4	248	21	AAV77989
19	137	88.4	271	7	AAV60661
20	137	88.4	271	11	AAV04212
21	137	88.4	271	11	AAV04217
22	135	87.1	259	21	AAV58135
23	133	85.8	248	7	AAV60437
24	133	85.8	256	11	AAV04210
25	108	69.7	375	21	AAV77990
26	95	61.3	351	18	AAV18780
27	95	61.3	351	18	AAV13672
28	95	61.3	371	16	AAV75642
29	84	54.2	249	21	AAV15401
30	84	54.2	335	12	AAV12222
31	81	52.3	27	21	AAV77987
32	81	52.3	27	22	AAV63348
33	81	52.3	228	20	AAV03769
34	81	52.3	248	14	AAV45005
35	81	52.3	248	20	AAV29485
36	81	52.3	248	22	AAV36578
37	81	52.3	277	20	AAV41698
38	81	52.3	277	20	AAV25518
39	81	52.3	277	21	AAV44254
40	81	52.3	277	22	AAV28073
41	79.5	51.3	484	20	AAV99892
42	79.5	51.3	484	21	AAV81934
43	79.5	51.3	830	16	AAV65216
44	79	51.0	229	18	AAV15251
45	79	51.0	270	18	AAV15252

ALIGNMENTS

RESULT 1

AAV25519

ID AAY25519 standard: Protein; 27 AA.

XX AC AAY25519;

DT 30-SEP-1999 (first entry)

XX Human collectin protein consensus fragment.

XX Collectin: human; antibacterial; antiviral; treatment; infection.

OS Synthetic.

OS Homo sapiens.

XX WO9337767-A1.

XX 29-JUL-1999.

XX 24-JUL-1998; 98WO-JP03328.

XX 23-JAN-1998; 98JP-0011281.

XX (FUSO) FUSO PHARM IND LTD.

XX Wakamiya N;

XX WPI; 1999-458691/38.

XX New collectin protein of human origin and DNA encoding it

XX Claim 3; Page 44; 58pp; Japanese.

XX This invention describes the isolation and characterisation of a novel

```
CC human collectin protein and its encoding polynucleotide. The human
CC collectin exhibits antibacterial and antiviral activity and can be used
CC as an agent for the treatment of human bacterial and viral infections.
CC This sequence represents a human collectin protein consensus sequence.
XX
XX
SQ Sequence 27 AA;

Query Match 100.0%; Score 155; DB 20; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLSRLAICEF 27
Db 1 ekcvemytdgkwnrndncqlsrlaicef 27

RESULT 2
AAY77986
ID AAY77986 standard; peptide; 27 AA.
XX
AC AAY77986;
XX
XX 20-JUN-2000 (first entry)
XX
XX Modified consensus sequence of collectin.
XX
XX Collectin; human; antibacterial; antiviral.
XX
XX Homo sapiens.
XX
XX WO200011161-A1.
XX
XX 02-MAR-2000.
XX
XX 24-AUG-1999; 99WO-JP04552.
XX
XX 24-AUG-1998; 98JP-0237611.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2000-224696/19.
XX
XX New collectin of human origin having antibacterial and antiviral
XX activity, and gene encoding it useful for production of transgenic
XX animals and of antibodies for screening potential drug molecules -
XX
XX Example 1; Page 31; 106pp; Japanese.
XX
XX The invention relates to polynucleotides encoding a new collectin of
XX human origin. The collectin can be used as an antibacterial and antiviral
XX agent and for screening potential drug molecules. The new collectin can
XX be produced by standard recombinant methodology. The present sequence
XX represents a modified consensus sequence of collectins hybridisable with
XX the novel human collectin.
XX
XX
SQ Sequence 27 AA;

Query Match 100.0%; Score 155; DB 21; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLSRLAICEF 27
Db 1 ekcvemytdgkwnrndncqlsrlaicef 27

RESULT 3
AAG63349
ID AAG63349 standard; protein; 27 AA.
```

```
XX
AC AAG63349;
XX
DI 15-OCT-2001 (first entry)
XX
DE Consensus sequence of a collectin.
XX
XX Human; scavenger receptor; SRC1-P1; macrophage; basal immunity;
XX arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
XX low density lipoprotein; collectin.
XX
XX Synthetic.
XX
XX WO200159107-A1.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
XX
XX 14-FEB-2000; 2000JP-0035155.
XX
XX 10-OCT-2000; 2000JP-0309068.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2001-497076/54.
XX
XX New Scavenger receptor proteins SRC1-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation -
XX
XX Example 1; Page 47; 118pp; Japanese.
XX
XX The present sequence represents a collectin consensus sequence. The
XX specification describes human scavenger receptors designated
XX SRC1-P1. The SRC1-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX
SQ Sequence 27 AA;

Query Match 100.0%; Score 155; DB 22; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLSRLAICEF 27
Db 1 ekcvemytdgkwnrndncqlsrlaicef 27

RESULT 4
AAP60439
ID AAP60439 standard; protein; 117 AA.
XX
XX AAP60439;
XX
XX 01-JAN-1980 (first entry)
XX
XX 3' Terminal portion of human alveolar surfactant protein (ASP).
XX
XX Alveolar surfactant protein; ASP; respiratory distress syndrome;
XX respiratory disease; pneumonia; bronchitis; ss.
XX
XX Homo sapiens.
XX
XX WO9603408-A.
XX
XX
```

```
XX PD 19-JUN-1986.
XX PF 10-DEC-1985; 85WO-US02445.
XX PR 11-DEC-1984; 84US-0680358.
XX PR 13-DEC-1985; 85US-0808843.
XX PR 29-JAN-1987; 87US-0008453.
XX PA (CALB-) CALIF BIOTECHNOL IN.
XX PI Schilling JW, White RT, Cordell B, Benson BJ;
DR WPI: 1986-169372/26.
DR N-PSDB; AAN60377.
XX Recombinant alveolar surfactant protein - for treatment of
PT respiratory distress syndrome and respiratory diseases
XX PS Disclosure; Fig. 5; 77pp; English.
XX CC The ASP can be used for the treatment of respiratory distress
CC syndrome in infants or adults and respiratory diseases such as
CC pneumonia and bronchitis. Preferably, the 32k component in
CC combination with the 10 k component of ASP is combined with
CC natural or synthetic lipids to construct a surfactant complex. See
CC also AAN60375-6, AAN60378-87 and AAP60437-8, AAP60440-42 and
CC AAP60445-48.
XX SQ Sequence 117 AA;

Query Match 88.4%; Score 137; DB 7; Length 117;
Best Local Similarity 85.2%; Pred. No. 1e-11;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 91 eqcvemytdgqwndrnclysrliceef 117
:|||||:|||||:|||||

RESULT 5
AAP60665
ID AAP60665 standard; Protein; 248 AA.
XX AC AAP60665;
XX DT 31-JUL-1991 (first entry)
XX DE Sequence of human alveolar surfactant protein (hASP)
DE on pMT(E):HS and pASPC-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;

Query Match 88.4%; Score 137; DB 7; Length 117;
Best Local Similarity 85.2%; Pred. No. 1e-11;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 91 eqcvemytdgqwndrnclysrliceef 117
:|||||:|||||:|||||

RESULT 5
AAP60665
ID AAP60665 standard; Protein; 248 AA.
XX AC AAP60665;
XX DT 31-JUL-1991 (first entry)
XX DE Sequence of human alveolar surfactant protein (hASP)
DE on pMT(E):HS and pASPC-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;
```

```
XX WPI: 1986-238888/36.
DR N-PSDB; AAN60571.
XX Regulatable expression systems - contg. human metallo:thionein-II
PT promoter
XX Example; Fig 5; 94pp; English.
XX A regulatable expression system for a coding sequence is claimed.
CC The system can process genomic as well as intronless DNA.
XX SQ Sequence 248 AA;

Query Match 88.4%; Score 137; DB 7; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdgqwndrnclysrliceef 248
:|||||:|||||:|||||

RESULT 6
AAP60666
ID AAP60666 standard; Protein; 248 AA.
XX AC AAP60666;
XX DT 31-JUL-1991 (first entry)
XX DE Genomic sequence of human alveolar surfactant protein (hASP)
DE encoded by genomic DNA, used to obtain pASPCg-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;

Query Match 88.4%; Score 137; DB 7; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdgqwndrnclysrliceef 248
:|||||:|||||:|||||

RESULT 6
AAP60666
ID AAP60666 standard; Protein; 248 AA.
XX AC AAP60666;
XX DT 31-JUL-1991 (first entry)
XX DE Genomic sequence of human alveolar surfactant protein (hASP)
DE encoded by genomic DNA, used to obtain pASPCg-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;

Query Match 88.4%; Score 137; DB 7; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdgqwndrnclysrliceef 248
:|||||:|||||:|||||

RESULT 6
AAP60666
ID AAP60666 standard; Protein; 248 AA.
XX AC AAP60666;
XX DT 31-JUL-1991 (first entry)
XX DE Genomic sequence of human alveolar surfactant protein (hASP)
DE encoded by genomic DNA, used to obtain pASPCg-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;
```

Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWDRNCQSLRLAICEF 27
 Db 222 EKCVMYTDGKWDRNCQSLRLAICEF 248

RESULT 7

AAP60441
 ID AAP60441 standard; protein; 248 AA.
 XX
 XX AC AAP60441;
 XX
 XX 01-JAN-1980 (first entry)
 XX
 XX Plasmid pASpc-SV(10) encoding human alveolar surfactant
 DE protein.
 XX
 XX Alveolar surfactant protein; ASP; canine; plasmid pASpc-SV(10);
 KW respiratory distress syndrome; respiratory disease; pneumonia;
 KW bronchitis; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO8603408-A.
 XX
 XX 19-JUN-1986.
 XX
 XX 10-DEC-1985; 85WO-US02445.
 XX
 XX 11-DEC-1984; 84US-0680358.
 PR 13-DEC-1985; 85US-0808843.
 PR 29-JAN-1987; 87US-0008453.
 XX
 XX (CALB-) CALIF BIOTECHNOL IN.
 XX
 XX Schilling JW, White RT, Cordell B, Benson BJ;
 PI WPI: 1986-169372/26.
 DR N-PSDB; AAN60379.
 XX
 XX Recombinant alveolar surfactant protein - for treatment of
 PT respiratory distress syndrome and respiratory diseases
 PT
 PS Disclosure; Fig. 7; 77pp; English.
 XX
 XX The ASP can be used for the treatment of respiratory distress
 CC syndrome in infants or adults and respiratory diseases such as
 CC pneumonia and bronchitis. This plasmid is used for the expression
 CC of human ASP in mammalian cell culture, e.g CHO cell culture. See
 CC also AAN60375-78, AAN60380-87 and AAP60437-40, AAP60442-49.
 XX
 XX Sequence 248 AA;
 XX
 XX Schilling JW, White RT, Cordell B, Benson BJ;
 PI WPI: 1986-169372/26.
 DR N-PSDB; AAN60379.
 XX
 XX Recombinant alveolar surfactant protein - for treatment of
 PT respiratory distress syndrome and respiratory diseases
 PT
 PS Disclosure; Fig. 7; 77pp; English.

QY 1 EKCVMYTDGKWDRNCQSLRLAICEF 27
 Db 222 EKCVMYTDGKWDRNCQSLRLAICEF 248

Query Match 88.4%; Score 137; DB 7; Length 248;
 Best Local Similarity 85.2%; Pred. No. 2.2e-11;
 Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWDRNCQSLRLAICEF 27
 Db 222 EKCVMYTDGKWDRNCQSLRLAICEF 248

RESULT 8

AAP60442
 ID AAP60442 standard; protein; 248 AA.
 XX
 XX AC AAP60442;
 XX
 XX 01-JAN-1980 (first entry)
 XX
 XX Plasmid pASpcq-SV(10) encoding human alveolar surfactant
 DE

DE

XX Alveolar surfactant protein; ASP; canine; plasmid pASpcq-SV(10);
 KW respiratory distress syndrome; respiratory disease; pneumonia;
 KW bronchitis; ss.
 XX

OS Homo sapiens.

XX WO8603408-A.

XX 19-JUN-1986.

XX 10-DEC-1985; 85WO-US02445.

XX 11-DEC-1984; 84US-0680358.

PR 13-DEC-1985; 85US-0808843.

PR 29-JAN-1987; 87US-0008453.

XX (CALB-) CALIF BIOTECHNOL IN.

XX Schilling JW, White RT, Cordell B, Benson BJ;

XX WPI: 1986-169372/26.

DR N-PSDB; AAN60380.

XX Recombinant alveolar surfactant protein - for treatment of

PT respiratory distress syndrome and respiratory diseases

PT

PS Disclosure; Fig. 8; 77pp; English.

XX The ASP can be used for the treatment of respiratory distress

CC syndrome in infants or adults and respiratory diseases such as

CC pneumonia and bronchitis. This plasmid is used for the expression

CC of human ASP in mammalian cell culture, e.g CHO cell culture. See

CC also AAN60375-79, AAN60381-87 and AAP60437-41, AAP60443-48.

XX Sequence 248 AA;

XX

QY 1 EKCVMYTDGKWDRNCQSLRLAICEF 27

Db 222 EKCVMYTDGKWDRNCQSLRLAICEF 248

RESULT 9

AAP70662
 ID AAP70662 standard; Protein; 248 AA.
 XX
 XX AAP70662;
 XX
 XX 29-APR-1991 (first entry)
 XX
 XX 35kd pulmonary surfactant protein.
 DE
 XX Hyaline membrane disease; respiratory distress syndrome; RDS.
 KW
 XX Homo sapiens.
 OS
 XX WO8702037-A.
 FN
 XX 09-APR-1987.
 PD
 XX 26-SEP-1986; 86WO-US02034.
 XX
 XX 26-SEP-1985; 85US-0781130.
 PR 15-AUG-1986; 86US-0897183.
 XX
 XX (GENE-) GENETICS INSTITUTE INC.
 PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.

XX
PI Tausch HW, Cobs KA, Steinbrink DR, Floros J, Phelps DS;
XX
DR WPI; 1987-108682/15.
DR N-PSDB; AAN71009.
XX
XX Pulmonary surfactant proteins - used for treating Hyaline Membrane
PT Disease or Respiratory Distress Syndrome.
XX
XX Claim 1; Page 33-34; 50pp; English.
XX
XX Gene product may be used in treatment of Hyaline Membrane Disease
CC and Respiratory Distress Syndrome (RDS) in both premature infants
CC and adults eg. cardio-pulmonary operations. The protein products
CC may also be used to raise diagnostic antibodies.
XX
XX Sequence 248 AA;
SQ

Query Match 88.4%; Score 137; DB 8; Length 248;
Best Local Similarity .85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdggwdrnclysrlicef 248
|:|||||:|||||:|||||:|||||

RESULT 10
AAP70663
ID AAP70663 standard; Protein; 248 AA.
XX
XX AAP70663;
AC
XX 29-APR-1991 (first entry)
DT
XX
XX 35kd pulmonary surfactant protein.
DE
XX
XX Hyaline membrane disease; respiratory distress syndrome; RDS.
KW
XX Homo sapiens.
OS
XX WO8702037-A.
PN
XX 09-APR-1987.
PD
XX 26-SEP-1986; 86WO-US02034.
PF
XX 26-SEP-1985; 85US-0781130.
PR
XX 15-AUG-1986; 86US-0897183.
PR
XX (GENE-) GENETICS INSTITUTE INC.
PA
XX (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.
PA
XX Tausch HW, Cobs KA, Steinbrink DR, Floros J, Phelps DS;
XX
XX WPI; 1987-108682/15.
DR
XX N-PSDB; AAN71010.
DR
XX
XX Pulmonary surfactant proteins - used for treating Hyaline Membrane
PT Disease or Respiratory Distress Syndrome.
PT
XX
XX Claim 1; Page 34A-B; 50pp; English.
PS
XX
XX Gene product may be used in treatment of Hyaline Membrane Disease
CC and Respiratory Distress Syndrome (RDS) in both premature infants
CC and adults eg. cardio-pulmonary operations. The protein products
CC may also be used to raise diagnostic antibodies.
CC
XX
XX Sequence 248 AA;
SQ

Query Match 88.4%; Score 137; DB 8; Length 248;

Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdggwdrnclysrlicef 248
|:|||||:|||||:|||||:|||||

RESULT 11
AAP80694
ID AAP80694 standard; protein; 248 AA.
XX
XX AAP80694;
AC
XX 04-DEC-1990 (first entry)
DT
XX
XX Sequence deduced from pHS10-5, encoding human 32K ASP.
DE
XX
XX Alveolar surfactant protein; ASP; respiratory distress syndrome;
KW pneumonia; bronchitis; 32K.
KW
XX Homo sapiens.
XX
XX WO8805820-A.
PN
XX 11-AUG-1988.
PD
XX
XX 15-JAN-1987; 87WO-US00092.
PF
XX
XX 29-JAN-1987; 87US-0008453.
PR
XX (CALB-) CALIF BIOTECHN INC.
PA
XX
XX Schilling JW, White RT, Cordell B, Benson BJ;
PI
XX WPI; 1988-124493/18.
DR
XX
XX Recombinant alveolar surfactant protein - used for treating
PT respiratory distress syndrome and related diseases e.g. pneumonia
PT and bronchitis.
PT
XX
XX Disclosure; ; pp; English.
PS
XX
XX The sequence was deduced from clone pHS10-5, isolated from a
CC human lung cDNA library. The protein is part of the alveolar
CC surfactant protein, high mol. wt, hydrophilic 32k gp. The protein
CC differs at position 50 from a previously published (WO8603408)
CC sequence, and differs at several places from two other sequences
CC determined by others. It is believed that the 32K human ASP
CC protein may be encoded by multiple genes. The recombinant protein
CC can be used for the treatment of respiratory disorders.
CC See also AAP82977-80, and AAP82982.
XX
XX Sequence 248 AA;
SQ

Query Match 88.4%; Score 137; DB 9; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdggwdrnclysrlicef 248
|:|||||:|||||:|||||:|||||

RESULT 12
AAP82980
ID AAP82980 standard; protein; 248 AA.
XX
XX AAP82980;
AC
XX
XX 04-DEC-1990 (first entry)
DT
XX

DE Sequence deduced from pHS10-4, encoding human 32K ASP.
XX
KW Alveolar surfactant protein; ASP; respiratory distress syndrome;
KW pneumonia; bronchitis; 32K.
XX
OS Homo sapiens.
XX
PN W0805820-A.
XX
PD 11-AUG-1988.
XX
XX 15-JAN-1987; 87WO-US00092.
XX
XX 29-JAN-1987; 87US-0008453.
XX
XX (CALB-) CALIF BIOTECH INC.
XX
XX Schilling JW, White RT, Cordell B, Benson BJ;
XX
XX WPI; 1988-124493/18.
XX
XX Recombinant alveolar surfactant protein - used for treating
PT respiratory distress syndrome and related diseases e.g. pneumonia
PT and bronchitis.
XX
XX Disclosure: ; pp; English.
XX
XX The sequence was deduced from clone pHS10-4, isolated from a
CC human lung cDNA library. The protein is part of the alveolar
CC surfactant protein, high mol. wt, hydrophilic 32K gp. The protein
CC differs at 7 positions from a previously published (W08603408)
CC sequence, and also at several places from two other sequences
CC determined by others. It is believed that the 32K human ASP
CC protein may be encoded by multiple genes. The recombinant protein
CC can be used for the treatment of respiratory disorders.
CC See also AAF82977-79, AAF82982 and AAF80694.
XX
XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 9; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvmytdgqwndrnclyslrticef 248

RESULT 13
AAR05092
ID AAR05092 standard; protein; 248 AA.
XX
XX AAR05092;
XX
XX 11-JUL-1990 (first entry)
XX
XX Product of vector PSP 35K-68-A encoding pulmonary surfactant protein.
DE
XX Respiratory distress syndrome; RDS; hyaline membrane disease;
KW pulmonary surfactant protein;
XX
XX Homo sapiens.
XX
XX US4882422-A.
XX
XX 21-NOV-1989.
XX
XX 24-SEP-1987; 87US-0100372.
XX
XX 26-SEP-1985; 85US-0791120.
PR 15-AUG-1986; 86US-0897183.
XX

PA (GENE-) GENETICS INSTITUTE.
XX
XX Taush HW, Jacobs KA, Steinbrink DR, Floros J, Phelps DS, Fritsh EF;
XX
XX WPI; 1990-036829/05.
DR N-PSDB; AAO02893.
XX
XX Purified human pulmonary surfactant protein -
PT useful for treating respiratory distress syndrome.
PT
XX
XX Disclosure; Table 2; 15pp; English.
XX
XX Protein is useful in treatment of respiratory distress syndrome (Hyaline
CC membrane disease) enhancing pulmonary surfactant activity.
CC Protein is encoded by the cDNA insert in vector PSP 35K-68-A
CC (ATCC 40244).
XX
XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 11; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvmytdgqwndrnclyslrticef 248

RESULT 14
AAR05091
ID AAR05091 standard; protein; 248 AA.
XX
XX AAR05091;
XX
XX 11-JUL-1990 (first entry)
XX
XX Vector PSP 35K-1A-10 gene product encoding pulmonary surfactant protein.
DE
XX Respiratory distress syndrome; RDS; hyaline membrane
KW disease; pulmonary surfactant protein;
XX
XX Homo sapiens.
XX
XX US4882422-A.
PN
XX 21-NOV-1989.
PD
XX 24-SEP-1987; 87US-0100372.
XX
XX 26-SEP-1985; 85US-0791120.
PR 15-AUG-1986; 86US-0897183.
XX
XX (GENE-) GENETICS INSTITUTE.
XX
XX Taush HW, Jacobs KA, Steinbrink DR, Floros J, Phelps DS, Fritsh EF;
XX
XX WPI; 1990-036829/05.
DR N-PSDB; AAO02892.
XX
XX Purified human pulmonary surfactant protein -
PT useful for treating respiratory distress syndrome.
PT
XX
XX Disclosure; Table 1; 15pp; English.
XX
XX Protein is useful in treatment of respiratory distress syndrome (Hyaline
CC membrane disease) enhancing pulmonary surfactant activity.
CC Protein is encoded by the cDNA insert in vector PSP 35K-1A-10
CC (ATCC 40243).
XX
XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 11; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVENYTDGKWNDRNCLOSRLAICEF 27
|:|||||:||||| ||| ||||
Db 222 egcvemytdgqwndrnclysrliceef 248

RESULT 15
AAR04215
ID AAR04215 standard; protein; 248 AA.
XX AC AAR04215;
XX DT 12-SEP-1989 (first entry)
XX DE Human 32K ASP encoded by pASpc-SV(10).
XX KW Mammalian 32K alveolar surfactant protein; respiratory distress syndrome;
XX LW lungs.
XX OS synthetic.
XX PN US4912038-A.
XX PD 27-MAR-1990.
XX PF 13-DEC-1985; 85US-0808843.
XX PR 11-DEC-1984; 84US-0680358.
XX PA (CALB-) CALIFORNIA BIOTECHNOLOGY INC.
XX PI Schilling JW, White RT, Cordell B, Benson BJ;
XX DR WPI; 1990-139399/18.
XX DR N-PSDB; AAQ04080.
XX PT Recombinant vector contg. DNA encoding 32K alveolar surfactant
XX PT protein used for treatment of eg respiratory distress syndrome
XX PS Claim 1; Fig 7; 70pp; English.
XX CC This sequence is the human 32K alveolar surfactant protein which is
XX CC encoded by the expression vector pASpc-SV(10) which is used to transform
XX CC mammalian host cells (eg CHO cells) resulting in efficient expression
XX CC of the protein.
XX CC See also AAR04210-12, AAR04214 and R04216-17.
XX SQ Sequence 248 AA;

Query Match 88.4%; Score 137; DB 11; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVENYTDGKWNDRNCLOSRLAICEF 27
|:|||||:||||| ||| ||||
Db 222 egcvemytdgqwndrnclysrliceef 248

Search completed: July 3, 2002, 12:34:20
Job time: 374 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:47 ; Search time 20.65 Seconds
(without alignments)
31.937 Million cell updates/sec

Title: US-09-600-932-3
Perfect score: 155
Sequence: 1 EKCVMYTDGKWNDRNCLOSLAICEF 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425534 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	84.5	111	6	Patent No. 5514582
2	95	61.3	161	3	Sequence 6, Appli
3	95	61.3	351	3	US-09-011-735-6
4	81.5	52.6	830	5	PCT-US91-05059-2
5	81.5	52.6	830	6	5378464-2
6	81	52.3	248	1	US-09-198-603C-2
7	80.5	51.9	119	1	US-08-340-539A-16
8	79.5	51.3	117	1	US-08-274-661B-39
9	79.5	51.3	119	1	US-08-340-539A-17
10	79.5	51.3	484	2	US-08-252-493C-9
11	79.5	51.3	484	3	US-09-276-197-9
12	79.5	51.3	830	1	US-08-110-158-4
13	79	51.0	287	4	US-09-111-470-6
14	74	47.7	114	6	5514582-8
15	72.5	46.8	119	1	US-08-340-539A-15
16	72.5	46.8	120	1	US-08-274-661B-36
17	72.5	46.8	120	1	US-08-274-661B-37
18	72.5	46.8	610	1	US-08-365-470-3
19	72.5	46.8	610	3	US-09-209-668-19
20	72.5	46.8	610	4	US-09-009-490A-89
21	72.5	46.8	610	6	5217870-2
22	72	46.5	129	3	US-08-722-126A-10
23	72	46.5	129	5	PCT-US95-04258-10
24	70.5	45.5	119	1	US-08-340-539A-14
25	69.5	44.8	119	1	US-08-340-539A-18
26	68.5	44.2	117	6	5514582-7
27	68.5	44.2	119	1	US-08-340-539A-12

28	68.5	44.2	119	1	US-08-340-539A-13
29	68.5	44.2	372	2	US-08-513-278-4
30	68.5	44.2	372	6	5514582-4
31	67.5	43.5	119	1	US-08-340-539A-19
32	66.5	42.9	117	1	US-08-274-661B-38
33	66.5	42.9	372	2	US-08-513-278-2
34	66.5	42.9	372	6	5514582-2
35	66.5	42.9	385	1	US-08-340-539A-2
36	66.5	42.9	385	2	US-08-461-592B-2
37	66.5	42.9	1455	3	US-08-840-062-5
38	66	42.6	107	6	5514582-17
39	66	42.6	108	6	5514582-16
40	62.5	40.3	1479	3	US-08-840-062-2
41	62.5	40.3	1479	3	US-08-840-062-4
42	62	40.0	114	6	5514582-13
43	62	40.0	114	6	5514582-14
44	62	40.0	273	4	US-09-111-470-10
45	62	40.0	292	2	US-08-688-342-4

ALIGNMENTS

RESULT 1
5514582-20
; Patent No. 5514582
; APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
; IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 43
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,670
; FILING DATE: 21-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 986,931
; FILING DATE: 08-DEC-1992
; APPLICATION NUMBER: 808,122
; FILING DATE: 16-DEC-1991
; APPLICATION NUMBER: 440,625
; FILING DATE: 22-NOV-1989
; APPLICATION NUMBER: 315,015
; FILING DATE: 23-FEB-1989
; SEQ ID NO:20:
; LENGTH: 111
5514582-20

Query Match 84.5%; Score 131; DB 6; Length 111;
Best Local Similarity 88.0%; Pred. No. 7.1e-12;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAIC 25
|||||
Db 87 EKCVMYTDGKWNDRNCLOSLAIC 111

RESULT 2
US-09-011-735-6
; Sequence 6, Application US/09011735B
; Patent No. 6110708
; GENERAL INFORMATION:
; APPLICANT: Wakamiya, No. 6110708utaka
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof
; FILE REFERENCE: 19036/34548
; CURRENT APPLICATION NUMBER: US/09/011,735B
; CURRENT FILING DATE: 1998-05-22
; EARLIER APPLICATION NUMBER: JP 7-209698
; EARLIER FILING DATE: 1995-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 161
; TYPE: PRT

ORGANISM: Bovine
FEATURE:
NAME/KEY: VARIANT
LOCATION: (2)
OTHER INFORMATION: Xaa is a protein-constituting amino acid
FEATURE:
NAME/KEY: VARIANT
LOCATION: (3)
OTHER INFORMATION: Xaa is a protein-constituting amino acid
FEATURE:
NAME/KEY: VARIANT
LOCATION: (5)
OTHER INFORMATION: Xaa is a protein-constituting amino acid
FEATURE:
NAME/KEY: VARIANT
LOCATION: (6)
OTHER INFORMATION: Xaa is a protein-constituting amino acid
US-09-011-735-6

Query Match 61.3%; Score 95; DB 3; Length 161;
Best Local Similarity 59.3%; Pred. No. 1.5e-06;
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRCNLCQSLAICEF 27
DB 135 ENCVEIFPDGKNDVPCSKQLLVICF 161

RESULT 3

US-09-011-735-1
Sequence 1, Application US/09011735B
Patent No. 6110708
GENERAL INFORMATION:
APPLICANT: Wakamaya, No. 6110708utaka
TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof
FILE REFERENCE: 19036/34548
CURRENT APPLICATION NUMBER: US/09/011,735B
CURRENT FILING DATE: 1998-05-22
EARLIER APPLICATION NUMBER: JP 7-209698
EARLIER FILING DATE: 1995-08-17
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 351
TYPE: PRT
ORGANISM: Bovine
US-09-011-735-1

Query Match 61.3%; Score 95; DB 3; Length 351;
Best Local Similarity 59.3%; Pred. No. 3.5e-06;
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRCNLCQSLAICEF 27
DB 325 ENCVEIFPDGKNDVPCSKQLLVICF 351

RESULT 4

PCT-US91-05059-2
Sequence 2, Application PC/TUS9105059
GENERAL INFORMATION:
APPLICANT: Regents of the Board of the, University of
APPLICANT: Oklahoma
TITLE OF INVENTION: Functionally Active Selectin-Derived
Peptides
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street, Suite 3100
CITY: Atlanta
STATE: Georgia

COUNTRY: US
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/05059
FILING DATE: 19910717
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/320408
FILING DATE: 08-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/554199
FILING DATE: 17-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Padst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF110CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEX: 404-572-6555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 830 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: YES
ANTI-SENSE: YES
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Blood
CELL TYPE: Endothelial
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 4..25
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 60..158
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 131..150
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 163..174
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 168..183
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 185..194
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 200..244
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 213..226
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 230..257
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 262..306
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 275..288
FEATURE:
NAME/KEY: Disulfide-bond

Db 129 EDCVEIYIKSPAGKWNDEMCLKKKALC 158

RESULT 6

US-09-198-603C-2

; Sequence 2, Application US/09198603C

; Patent No. 6337193

; GENERAL INFORMATION:

; APPLICANT: TULLY, Raymond E.

; APPLICANT: CALTAGIRONE, G. Thomas

; APPLICANT: MOYER, Shawn S.

; APPLICANT: KONNING, Michael T.

; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC

; FILE REFERENCE: YEAST

; CURRENT APPLICATION NUMBER: A7290

; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 248

; TYPE: PRT

; ORGANISM: Human

US-09-198-603C-2

Query Match

Best Local Similarity 52.3%; Score 81; DB 4; Length 248;

Matches 14; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27

|||||:|||||:|||||:|||||

Db 220 EDCVLLKNGQNDVPCSTSLAYCEF 246

RESULT 7

US-08-340-539A-16

; Sequence 16, Application US/08340539A

; Patent No. 5808025

; GENERAL INFORMATION:

; APPLICANT: Tedder, Thomas F.

; APPLICANT: Kansas, Geoffrey S.

; TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FISH & NEAVE

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/340,539A

; FILING DATE: 16-NOV-1994

; CLASSIFICATION: 514

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/008,459

; FILING DATE: 25-JAN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Gunnison, Jane

; REGISTRATION NUMBER: 38,479

; REFERENCE/DOCKET NUMBER: CG-104 CON

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-596-9000

; TELEFAX: 212-596-9090

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 119 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-340-539A-16

Query Match

Best Local Similarity 51.9%; Score 80.5; DB 1; Length 119;

Matches 14; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 EKCVMY-----TDGKWNDRNCLOSLAIC 25

|||||:|||||:|||||:|||||

Db 88 EDCVEIYIKREKDSGKWNDEKCTKOKLALC 117

RESULT 8

US-08-274-661B-39

; Sequence 39, Application US/08274661B

; Patent No. 5593882

; GENERAL INFORMATION:

; APPLICANT: Erbe, David V.

; APPLICANT: Lasky, Laurence A.

; APPLICANT: Presta, Leonard G.

; TITLE OF INVENTION: Selectin Variants

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Winpatin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/274,661B

; FILING DATE: 13-Jul-1994

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 07/956701

; FILING DATE: 10/01/1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: 761P1C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-3216

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 39:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 117 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

US-08-274-661B-39

Query Match

Best Local Similarity 51.3%; Score 79.5; DB 1; Length 117;

Matches 14; Conservative 5; Mismatches 6; Indels 5; Gaps 1;

QY 1 EKCVMY-----TDGKWNDRNCLOSLAIC 25

|||||:|||||:|||||:|||||

Db 88 EDCVEIYIKSPAGKWNDEKCLKKHALC 117

RESULT 9

US-08-340-539A-17


```
; Sequence 17, Application US/08340539A
; Patent No. 5808025
; GENERAL INFORMATION:
; APPLICANT: Tedder, Thomas F.
; APPLICANT: Kansas, Geoffrey S.
; TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMUL/PANEOUS
; TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,539A
; FILING DATE: 16-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/008,459
; FILING DATE: 25-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gunnison, Jane
; REGISTRATION NUMBER: 38,479
; REFERENCE/DOCKET NUMBER: CG-104 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-340-539A-17

Query Match 51.3%; Score 79.5; DB 1; Length 119;
Best Local Similarity 46.7%; Pred. No. 0.00019;
Matches 14; Conservative 5; Mismatches 6; Indels 5; Gaps 1;

Qy 1 EKCVMY-----TDGKWNDRNCQSRLAIC 25
| ||||| | : |||
Db 88 EDCVEIYKSPAGKWNDEHCLKKKHALC 117

RESULT 10
US-08-252-493C-9
; Sequence 9, Application US/08252493C
; Patent No. 5891645
; GENERAL INFORMATION:
; APPLICANT: Rollins, Scott
; APPLICANT: Rother, Russell P.
; APPLICANT: Evans, Mark J.
; APPLICANT: Matis, Louis A.
; TITLE OF INVENTION: PORCINE E-SELECTIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seth A. Fidel
; STREET: 25 Science Park, Box 15
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06511
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 750 Kb storage
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,197
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/252,493
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
```

```
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,493C
; FILING DATE: June 1, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 776-1790
; TELEFAX: (203) 772-3655
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 amino acids
; TYPE: amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: predicted amino acid sequence of
; DESCRIPTION: Porcine E-selectin
US-08-252-493C-9

Query Match 51.3%; Score 79.5; DB 2; Length 484;
Best Local Similarity 46.7%; Pred. No. 0.00083;
Matches 14; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

Qy 1 EKCVMY-----TDGKWNDRNCQSRLAIC 25
| ||||| | : |||
Db 110 EDCVEIYKRDGSKGNDRCKSKKLAIC 139

RESULT 11
US-09-276-197-9
; Sequence 9, Application US/09276197
; Patent No. 6040428
; GENERAL INFORMATION:
; APPLICANT: Rollins, Scott
; APPLICANT: Rother, Russell P.
; APPLICANT: Evans, Mark J.
; APPLICANT: Matis, Louis A.
; TITLE OF INVENTION: PORCINE E-SELECTIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seth A. Fidel
; STREET: 25 Science Park, Box 15
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06511
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 750 Kb storage
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,197
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/252,493
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
```


Search completed: July 3, 2002, 12:34:48
Job time: 192 sec

```
QY      1  EKCVEMYTDGKWNDRNCLQSLAIC 25
      ||||: ||:|||| ||| :|
Db     90  EDCVEVQPDGRWNDDFCLOVYRWVC 114
```

Query Match 46.8%; Score 72.5; DB 1; Length 119;
Best Local Similarity 43.3%; Pred. No. 0.0019;
Matches 13; Conservative 4; Mismatches 8; Indels 5; Gaps 1;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:40:21 ; Search time 42.99 Seconds
(without alignments)
108.650 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155
Sequence: 1 EKCVMTDQKWNDRNCQSLRAICEF 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_todent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140	90.3	248	6 Q9N0G1	Q9n0g1 equus cabal
2	140	90.3	248	6 Q95L88	Q95l88 equus cabal
3	140	90.3	248	11 Q9CQI1	Q9cqil mus musculu
4	139	89.7	116	6 Q9NIX3	Q9nlx3 sus scrofa
5	138	89.0	248	6 Q9T0G6	Q9tt06 ovis aries
6	110	71.0	222	13 Q90XB2	Q90xb2 gallus gall
7	99	63.9	378	6 Q9NIX4	Q9nlx4 sus scrofa
8	95	61.3	335	6 Q97748	Q97748 bos taurus
9	85	54.8	236	6 Q28518	Q28518 macaca mula
10	85	54.8	240	6 Q9XSW3	Q9xsw3 sus scrofa
11	83.5	53.9	610	6 Q95LG1	Q95lg1 equus cabal
12	82.5	53.2	484	6 Q95LG2	Q95lg2 ovis aries
13	81	52.3	248	4 Q96TF9	Q96tf9 homo sapien
14	81	52.3	248	4 Q96TF8	Q96tf8 homo sapien
15	81	52.3	248	4 Q96TF7	Q96tf7 homo sapien
16	81	52.3	248	4 Q96KE4	Q96ke4 homo sapien

17	81	52.3	277	4 Q9Y6Z7	Q9y6z7 homo sapien
18	80	51.6	245	6 Q28517	Q28517 macaca mula
19	79.5	51.3	482	6 Q28982	Q28982 sus scrofa
20	79.5	51.3	485	6 Q95LG3	Q95lg3 odocoileus
21	79.5	51.3	616	4 Q95509	Q95509 homo sapien
22	79.5	51.3	740	4 Q95507	Q95507 homo sapien
23	79.5	51.3	740	4 Q95508	Q95508 homo sapien
24	79	51.0	287	4 Q03969	Q03969 homo sapien
25	79	51.0	292	4 Q00448	Q00448 homo sapien
26	77.5	50.0	609	6 Q9GLF0	Q9glf0 canis famil
27	77	49.7	224	11 Q9Z294	Q9z294 rattus sp.
28	74.5	48.1	646	6 Q29097	Q29097 sus scrofa
29	72.5	46.8	376	6 Q28629	Q28629 oryctolagus
30	71	45.8	271	4 Q9BWE8	Q9bwp8 homo sapien
31	69	44.5	272	11 Q9DC75	Q9dc75 mus musculu
32	68.5	44.2	372	11 Q63762	Q63762 rattus norv
33	67.5	43.5	649	6 Q28657	Q28657 oryctolagus
34	67.5	43.5	754	6 Q28290	Q28290 canis famil
35	66.5	42.9	385	4 Q9CJ43	Q9cj43 homo sapien
36	66.5	42.9	1456	11 Q61830	Q61830 mus musculu
37	65	41.9	196	11 Q9EPW4	Q9epw4 mus musculu
38	65	41.9	197	4 Q75596	Q75596 homo sapien
39	65	41.9	284	11 Q91V84	Q91v84 mus musculu
40	62.5	40.3	1479	4 Q9UBG0	Q9ubg0 homo sapien
41	62.5	40.3	1479	4 Q9Y5P9	Q9y5p9 homo sapien
42	62.5	40.3	1479	11 Q64449	Q64449 mus musculu
43	62	40.0	292	4 Q14538	Q14538 homo sapien
44	61.5	39.7	238	13 Q57451	Q57451 gallus gall
45	61.5	39.7	254	13 Q98TA4	Q98ta4 gallus gall

ALIGNMENTS

```
RESULT 1
Q9N0G1
ID Q9N0G1 PRELIMINARY; PRT; 248 AA.
AC Q9N0G1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PULMONARY SURFACTANT PROTEIN A.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=THOROUGHRED; TISSUE=LUNG;
RA Hobo S.;
RT "Molecular cloning of equine pulmonary surfactant proteins.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB015963; BAA97976.1; -.
DR HSSP: P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 248 AA; 26000 MW; BBE12EFB05C2B8D1 CRC64;
```

Query Match 90.3%; Score 140; DB 6; Length 248;
Best Local Similarity 88.9%; Pred. No. 1e-13;
Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVMTDQKWNDRNCQSLRAICEF 27

|||||
Db 222 EKCVMTDQGNDRSCQYRLAICEF 248

RESULT 2

Q95L88
ID Q95L88 PRELIMINARY; PRT; 248 AA.
AC Q95L88;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN A.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Weber B.I.L., Hospes R., Gortner L.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF400580; AAL07690.1; -;
SQ SEQUENCE 248 AA; 26047 MW; B71133E005C9A5C1 CRC64;

Query Match 90.3%; Score 140; DB 6; Length 248;
Best Local Similarity 88.9%; Pred. No. 1e-13;
Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKNDKNCLOSLRAICEF 27
Db 222 EKCVEMTDGKNDKNCLOSLRAICEF 248
|||||
|||||

RESULT 3
Q9CQI1 PRELIMINARY; PRT; 248 AA.
ID Q9CQI1;
AC Q9CQI1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SURFACTANT ASSOCIATED PROTEIN A.
GN SFTPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO, AND LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK011333; BAB27551.1; -;
DR EMBL: AK004620; BAB23416.1; -;
DR EMBL: AK004788; BAB23565.1; -;
DR HSSP: P35247; I808.
DR MGD: MGI:109518; Sftpa.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; Lectin_c.
DR Pfam: PF00059; lectin_c; 1.

SMART; SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 248 AA; 26183 MW; 8A5670CFAD3EB9B6 CRC64;

Query Match 90.3%; Score 140; DB 11; Length 248;
Best Local Similarity 88.9%; Pred. No. 1e-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKNDKNCLOSLRAICEF 27
Db 222 EKCVEMTDGKNDKNCLOSLRAICEF 248
|||||
|||||

RESULT 4
Q9NIX3 PRELIMINARY; PRT; 116 AA.
ID Q9NIX3;
AC Q9NIX3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN A (FRAGMENT).
GN SFTPA.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20109098; PubMed=10640760;
RA van Eijk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.";
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF133668; AAF28384.1; -;
DR HSSP: P35247; I808.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SMC0034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
FT NON_TER 1
SQ SEQUENCE 116 AA; 12946 MW; 6601D39320760C74 CRC64;

Query Match 89.7%; Score 139; DB 6; Length 116;
Best Local Similarity 88.9%; Pred. No. 6.4e-14;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKNDKNCLOSLRAICEF 27
Db 90 EKCVEMTDGKNDKNCLOSLRAICEF 116
|||||
|||||

RESULT 5
Q9TT06 PRELIMINARY; PRT; 248 AA.
ID Q9TT06;
AC Q9TT06;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PULMONARY SURFACTANT PROTEIN A (PULMONARY SURFACTANT-ASSOCIATED
DE PROTEIN A).
GN SPAS OR SP-A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=20215262; PubMed=10749753;
RA Braems G.A., Yao L.-J., Inchley K., Brickenden A., Han V.K.M.,
RA Grolla A., Challis J.R.G., Rossmayer F.,
RT "Ovine surfactant protein cDNAs: use in studies on fetal lung growth
RT and maturation after prolonged hypoxemia.";
RL Am. J. Physiol. 278:L754-L764(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=20215263; PubMed=10749754;
RA Pietschmann S.M., Pison U.;
RT "cDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of
RT two different sequences for SP-B.";
RL Am. J. Physiol. 278:L765-L778(2000).
DR EMBL: AF211856; AAF18995.1; -
DR EMBL: AF076633; AAF31148.1; -
DR HSP: P19999; IYTI.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 248 AA; 26394 MW; D65E7293BBFF1FD9 CRC64;

Query Match 89.0%; Score 138; DB 6; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.1e-13;
Matches 23; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLQSLRAICEF 27
   |||||:|||||:|||||:|||||:|||||
Db 222 EKCVMYTDGKWNDRNCQLQSLRAICEF 248

RESULT 6
ID Q90XB2 PRELIMINARY; PRT; 222 AA.
AC Q90XB2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE SURFACTANT PROTEIN A PRECURSOR.
GN SP-A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Vitved L., Hansen S., Teisner B., Koch C., Juul-Nadsen H.,
RA Holmskov U., Salomonsen J., Skjoldt K.;
RT "The chicken homolog of surfactant protein A has only three collagen-
RT like Gly-Xaa-Yaa repeats and contain a new putative coil structure
RT between the collagen region and the alpha-helical coil-coil region.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF411083; AAK97540.1; -.
RW Signal.
FT SIGNAL 1 18 POTENTIAL.
SQ SEQUENCE 222 AA; 24721 MW; 97371A3C53303872 CRC64;

Query Match 71.0%; Score 110; DB 13; Length 222;
Best Local Similarity 66.7%; Pred. No. 3.9e-09;
Matches 18; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLQSLRAICEF 27
   |||||:|||||:|||||:|||||:|||||
Db 196 EKCVMYTDGKWNDRNCQLQSLRAICEF 222

```

```

RESULT 7
ID Q9N1X4 PRELIMINARY; PRT; 378 AA.
AC Q9N1X4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN D PRECURSOR.
GN SFTPD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20109098; PubMed=10640760;
RA van Eljk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.";
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF132496; AAF22145.2; -.
DR HSP: P35247; IB08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 378 LONG SURFACTANT PROTEIN D.
SQ SEQUENCE 378 AA; 37986 MW; 3504B8C1E56C341D CRC64;

Query Match 63.9%; Score 99; DB 6; Length 378;
Best Local Similarity 59.3%; Pred. No. 3.5e-07;
Matches 16; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLQSLRAICEF 27
   |||||:|||||:|||||:|||||:|||||
Db 352 ENCVEIFPNGKWNDRNCQLQSLRAICEF 378

RESULT 8
ID O97748 PRELIMINARY; PRT; 335 AA.
AC O97748;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONGLUTININ.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
RT primary site of synthesis of conglutinin in members of the Bovidae.";

```

Query Match 53.9%; Score 83.5; DB 6; Length 610;
Best Local Similarity 50.0%; Pred. No. 0.00014;
Matches 15; Conservative 4; Mismatches 6; Indels 5; Gaps 1;

Qy 1 EKCVMTD-----GKWNDRNCLOSRLAIC 25
| |||:| | ||| || : |||:|
Db 108 EDCVEIYIKRYKDAGKWNDRNCNKKKLALC 137

RESULT 12
Q95LG2
ID Q95LG2 PRELIMINARY; PRT; 484 AA.
AC Q95LG2
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE E-SELECTIN.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21421234; PubMed=11529941;
RA Hedges J.F., Demaula C.D., Moore B.D., McLaughlin B.E., Simon S.I.,
FA MacLachlan N.J.;
RT "Characterization of equine E-selectin."
RL Immunology 103:498-504(2001).
DR EMBL; AF307971; AAK48711.1; -.
KW Lectin; Selectin.
SQ SEQUENCE 484 AA; 52951 MW; 617930C1C2F47B44 CRC64;

Query Match 53.2%; Score 82.5; DB 6; Length 484;
Best Local Similarity 46.7%; Pred. No. 0.00016;
Matches 14; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

Qy 1 EKCVMTD-----TDGKWNDRNCLOSRLAIC 25
| |||:| | ||| || : |||:|
Db 109 EDCVEIYIKREKDSKWNDRNCNKKKFKALC 138

RESULT 13
Q96TF9
ID Q96TF9 PRELIMINARY; PRT; 248 AA.
AC Q96TF9;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MANNOSYL-BINDING LECTIN PRECURSOR.
GN MBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98414317; PubMed=9743385;
RA Madsen H.O., Satz L., Høgh B., Sveigaard A., Garred P.;
RT "Different molecular events result in low protein levels of mannan-
binding lectin in populations from South-East Africa and South
America."
RL J. Immunol. 161:3169-3175(1998).
DR EMBL; Y16578; CAB56045.1; -.
FT SIGNAL 1 20 POTENTIAL.
SQ SEQUENCE 248 AA; 26215 MW; 98934266AFB0F415 CRC64;

Query Match 52.3%; Score 81; DB 4; Length 248;
Best Local Similarity 51.9%; Pred. No. 0.00013;
Matches 14; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 EKCVMTDGGKWNDRNCLOSRLAICEF 27
| |||:| | ||| || : |||:|
Db 220 EDCVLLKNGQWNVPCSTSHLAVCEF 246

RESULT 14
Q96TF8
ID Q96TF8 PRELIMINARY; PRT; 248 AA.
AC Q96TF8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MANNOSYL-BINDING LECTIN PRECURSOR.
GN MBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98414317; PubMed=9743385;
RA Madsen H.O., Satz L., Høgh B., Sveigaard A., Garred P.;
RT "Different molecular events result in low protein levels of mannan-
binding lectin in populations from South-East Africa and South
America."
RL J. Immunol. 161:3169-3175(1998).
DR EMBL; Y16579; CAB56121.1; -.
FT SIGNAL 1 20 POTENTIAL.
SQ SEQUENCE 248 AA; 26201 MW; 6EF164F097CC3969 CRC64;

Query Match 52.3%; Score 81; DB 4; Length 248;
Best Local Similarity 51.9%; Pred. No. 0.00013;
Matches 14; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 EKCVMTDGGKWNDRNCLOSRLAICEF 27
| |||:| | ||| || : |||:|
Db 220 EDCVLLKNGQWNVPCSTSHLAVCEF 246

RESULT 15
Q96TF7
ID Q96TF7 PRELIMINARY; PRT; 248 AA.
AC Q96TF7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MANNOSYL-BINDING LECTIN PRECURSOR.
GN MBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98414317; PubMed=9743385;
RA Madsen H.O., Satz L., Høgh B., Sveigaard A., Garred P.;
RT "Different molecular events result in low protein levels of mannan-
binding lectin in populations from South-East Africa and South
America."
RL J. Immunol. 161:3169-3175(1998).
DR EMBL; Y16582; CAB56124.1; -.
FT SIGNAL 1 20 POTENTIAL.
SQ SEQUENCE 248 AA; 26090 MW; D4EECF85C6F48C4 CRC64;

Query Match 52.3%; Score 81; DB 4; Length 248;
Best Local Similarity 51.9%; Pred. No. 0.00013;
Matches 14; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 1 EKCVMTDGGKWNDRNCLOSRLAICEF 27
| |||:| | ||| || : |||:|
Db 220 EDCVLLKNGQWNVPCSTSHLAVCEF 246

Mon Jul 8 10:13:26 2002

us-09-600-932-3.rspt

Page 6

Search completed: July 3, 2002, 12:40:22
Job time: 361 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:35:22 ; Search time 28.2 Seconds
(without alignments)
92.000 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142	91.6	247	1 LNRBPS	pulmonary surfacta
2	140	90.3	248	2 A48853	pulmonary surfacta
3	137	88.4	248	1 LNHUPS	pulmonary surfacta
4	137	88.4	248	1 LNHUP6	pulmonary surfacta
5	137	88.4	248	1 LNHUP1	pulmonary surfacta
6	136	87.7	248	1 LNRTPS	pulmonary surfacta
7	134	86.5	248	2 I51921	pulmonary surfacta
8	133	85.8	248	1 LNDGFS	pulmonary surfacta
9	108	69.7	375	1 A45225	pulmonary surfacta
10	103	66.5	301	2 A53570	collectin-43 - bov
11	103	66.5	374	1 A42046	surfactant protein
12	100	64.5	369	2 S33603	surfactant protein
13	95	61.3	371	1 JN0450	conglutinin precu
14	95	61.3	371	2 I45878	conglutinin - bovi
15	87	56.1	244	1 LNRTRC	mannose-binding le
16	83	53.5	244	1 LNSMC	mannose-binding le
17	81	52.3	248	1 LNHUMC	mannose-binding le
18	80.5	51.9	485	2 S36772	E-selectin - pig
19	79.5	51.3	482	2 JC5092	E-selectin - bovin
20	79.5	51.3	830	2 A30359	P-selectin precurs
21	79	51.0	311	1 LNHU2A	asialoglycoprotein
22	77	49.7	238	1 LNRTHA	mannose-binding le
23	72.5	46.8	376	2 JC4832	L-selectin precurs
24	72.5	46.8	551	2 I46709	endothelial leukoc
25	72.5	46.8	610	2 A35046	E-selectin precurs
26	72.5	46.8	612	2 B42755	E-selectin precurs
27	72	46.5	301	2 S13165	asialoglycoprotein
28	71	45.8	239	1 LNSMA	mannose-binding le
29	70.5	45.5	370	2 S22124	L-selectin precurs

ALIGNMENTS

RESULT 1

LNRBPS

pulmonary surfactant protein A precursor - rabbit

N;Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

C;Accession: A29931

R;Bogdaram, V.; Qing, K.; Mendelson, C.R.

J. Biol. Chem. 263, 2939-2947, 1988

A;Title: The major apoprotein of rabbit pulmonary surfactant. Elucidation of primary

A;Reference number: A29931; MUID:88139348

A;Accession: A29931

A;Molecule type: mRNA

A;Residues: 1-247 <SOG>

A;Cross-references: GB:J03542; NID:g165705; PIDN:AAA31465.1; PTD:g165706

A;Note: 12-Pro was also found

A;Note: two species of mRNA, which appear to be transcribed from a single gene, could be identified by the amino end of the mature protein is blocked

C;Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower surface tension in the lungs. This protein is a stialoglycoprotein synthesized by alveolar type II cells.

C;Superfamily: mannose-binding lectin; C-type lectin homology

C;Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl

F;1-15/Domain: signal sequence #status predicted <SIG>

F;16-247/Product: pulmonary surfactant protein A #status predicted <MAT>

F;27-99/Region: collagen-like

F;126-245/Domain: C-type lectin homology <LCH>

F;16/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

F;206/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 91.68; Score 142; DB 1; Length 247;

Best Local Similarity 88.94; Pred. No. 3.5e-13;

Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27

Db 221 EKCVEMTDGKWNDRNCLOSLRAICEF 247

RESULT 2

A48853

pulmonary surfactant-associated protein SP-A - mouse

C;Species: Mus musculus (house mouse)

C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999

C;Accession: A48853

R;Korfhagen, T.R.; Bruno, M.D.; Glasser, S.W.; Cirraolo, P.J.; Whitsett, J.A.; Lattier

Am. J. Physiol. 263, L546-L554, 1992

A;Title: Murine pulmonary surfactant SP-A gene: cloning, sequence, and transcriptiona

A;Reference number: A48853; MUID:93072386

A;Contents: DBA/2J

A;Accession: A48853

```

A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-248 <KOR>
A:Cross-references: GB:S48768; NID:g260452; PIDN:AAE24274.1; PID:g260453
A:Note: sequence extracted from NCBI backbone (NCBIN:118740, NCBIPI:118741)
C:Superfamily: mannose-binding lectin; C-type lectin homology
F:127-246/Domain: C-type lectin homology <LCH>

Query Match      90.3%; Score 140; DB 2; Length 248;
Best Local Similarity 88.9%; Pred. No. 6.9e-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTGDKNDRNCLOSLRLAICEF 27
      ||||| ||||| ||||| ||||| |||||
Db 222 EKCVEMTGDKNDRNCLOSLRLAICEF 248

RESULT 3
LNHUP5
pulmonary surfactant protein A precursor (genomic clone) - human
N:Alternate names: alveolar proteinosis protein; pulmonary surfactant 32K apoprotein; pu
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
R:White, R.T.; Damm, D.; Miller, J.; Spratt, K.; Schilling, J.; Hawgood, S.; Benson, B.;
Nature 317, 361-363, 1985
A:Title: Isolation and characterization of the human pulmonary surfactant apoprotein gen
A:Reference number: A24622; MUID:86014366
A:Accession: A24622
A:Molecule type: DNA
A:Residues: 1-248 <WHI>
A:Cross-references: GB:M30838; NID:g190564; PIDN:AAA36510.1; PID:g190565
A:Note: the sequence in GenBank entry HUMPSAP, release 109.0, (PID:g190565) has the cod
A:Note: four nucleotide differences, producing amino acid differences at positions 45, 5
R:Haggsman, H.P.; White, R.T.; Schilling, J.; Lau, K.; Benson, B.J.; Golden, J.; Hawgood
Am. J. Physiol. 257, L421-L429, 1989
A:Title: Studies of the structure of lung surfactant protein SP-A.
A:Reference number: A43628; MUID:90119861
A:Accession: A43628
A:Molecule type: protein
A:Residues: 143-150;220-240;243-248 <HAA>
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t
C:Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells. It
pendent on the presence of calcium ions.
C:Genetics:
A:Gene: GDB:SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:119593; OMIM:178630
A:Map position: 10q22-10q23
A:Introns: 58/1; 98/1; 124/1
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: alveolar proteinosis; calcium; gaseous exchange; glycoprotein; hydroxylysine
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F:28-100/Domain: collagenous #status predicted <COL>
F:127-246/Domain: C-type lectin homology <LCH>
F:236/Disulfide bonds: interchain #status experimental
F:30,33,36,42,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status predict
F:51.88/Modified site: 5-hydroxylysine (Lys) #status predicted
F:155-246,224-238/Disulfide bonds: #status experimental
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      88.4%; Score 137; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 1.9e-12;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTGDKNDRNCLOSLRLAICEF 27
      ||||| ||||| ||||| ||||| |||||
Db 222 EKCVEMTGDKNDRNCLOSLRLAICEF 248

RESULT 4
LNHUP5
pulmonary surfactant protein A precursor (clone 1A) - human
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associat
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
R:Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman,
J. Biol. Chem. 261, 9029-9033, 1986
A:Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfa
A:Reference number: A25720; MUID:86250832
A:Accession: A25720
A:Molecule type: mRNA
A:Residues: 1-248 <FLO>
A:Cross-references: GB:M33686; NID:g190669; PIDN:AAA60211.1; PID:g190670
A:Note: part of the sequence was confirmed by protein sequencing
A:Note: the amino end of the mature protein, which was not identified, is partially a
A:Note: clones corresponding to two different proteins were sequenced. Cotranslationa
C:Genetics:
A:Gene: GDB:SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:119593; OMIM:178630
A:Map position: 10q22-10q23
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F:127-246/Domain: C-type lectin homology <LCH>
F:21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted
F:30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status p
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      88.4%; Score 137; DB 1; Length 248;

```

Best Local Similarity 85.2%; Pred. No. 1.9e-12;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
:|||||:|||||:|||||:|||||:|||||
Db 222 EQCVMTDGTWDRNCLOSLAICEF 248

RESULT 6
LNTPS
pulmonary surfactant protein A precursor - rat
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A29299; J50034; S23183
R:Sano, K.; Fisher, J.; Mason, R.J.; Kuroki, Y.; Schilling, J.; Benson, B.; Voelker, D.
Biochem. Biophys. Res. Commun. 144, 367-374, 1987
A:Title: Isolation and sequence of a cDNA clone for the rat pulmonary surfactant-associated
A:Reference number: A29299; MUID:87213191
A:Contents: 0.9 kb cDNA
A:Accession: A29299
A:Molecule type: mRNA
A:Residues: 1-248 <S>
A:Note: part of the sequence, including the amino end of the mature protein, was confirmed
R:Fisher, J.H.; Emrie, P.A.; Shannon, J.; Sano, K.; Hattler, B.; Mason, R.J.
Biochim. Biophys. Acta 950, 338-345, 1988
A:Title: Rat pulmonary surfactant protein A is expressed as two differently sized mRNA
A:Reference number: J50034; MUID:89000785
A:Contents: 1.6 kb cDNA
A:Accession: J50034
A:Molecule type: mRNA
A:Residues: 1-248 <PI>
A:Note: the codons given for residues 78, 84, and 180 are inconsistent with the authors
R:Lacaze-Masmon, T.; Fraslon, C.; Bourbon, J.; Raymondjean, M.; Kahn, A.
Eur. J. Biochem. 206, 613-623, 1992
A:Title: Characterization of the rat pulmonary surfactant protein A promoter.
A:Reference number: S23183; MUID:92298987
A:Accession: S23183
A:Molecule type: DNA
A:Residues: 1-32 <LAC>
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers the
pendent on the presence of calcium ions.
C:Comment: Two species of mRNA, which probably are transcribed from a single gene, have
C:Comment: Size heterogeneity of these proteins arises from posttranslational modification
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: alveolar proteinosis; calcium; gaseous exchange; glycoprotein; hydroxyproline
F:1-20/Domain: signal sequence #status predicted <Sig>
F:21-248/Product: pulmonary surfactant protein A #status experimental <MAT>
F:37-103/Region: collagen-like
F:127-246/Domain: C-type lectin homology <LCH>
F:21/Binding site: carbohydrate (Asn) (covalent) #status absent
F:30,33,36,42,54,57,63,67,70,76/Modified site: 4-hydroxyproline (Pro) #status experimental
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.7%; Score 136; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.6e-12;
Matches 23; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
:|||||:|||||:|||||:|||||:|||||
Db 222 EQCVMTDGTWDRNCLOSLAICEF 248

RESULT 7
I51921
pulmonary surfactant-associated protein A1 - human
N:Alternate names: SP-A1
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999
C:Accession: I51921
R:Katyal, S.L.; Singh, G.; Locker, J.

Am. J. Respir. Cell Mol. Biol. 6, 446-452, 1992
A:Title: Characterization of a second human pulmonary surfactant-associated protein
A:Reference number: I51921; MUID:92198680
A:Accession: I51921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-248 <RES>
A:Cross-references: GB:M69519; NID:g338048; PIDN:AAA60319.1; PID:g338049
C:Genetics:
A:Gene: GDB:SFTPA1; SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:I19593; OMIM:178630
A:Map position: 10q22-10q23
A:Introns: 58/1; 98/1; 124/1
C:Superfamily: mannose-binding lectin; C-type lectin homology
F:127-246/Domain: C-type lectin homology <LCH>

Query Match 86.5%; Score 134; DB 2; Length 248;
Best Local Similarity 81.5%; Pred. No. 5.1e-12;
Matches 22; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
:|||||:|||||:|||||:|||||:|||||
Db 222 EQCVMTDGTWDRNCLOSLAICEF 248

RESULT 8
LNDGFS
pulmonary surfactant protein A precursor - dog
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C:Species: Canis lupus familiaris (dog)
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A5296; A61227; A60142
R:Benson, B.; Hawgood, S.; Schilling, J.; Clements, J.; Damm, D.; Cordell, B.; White,
Proc. Natl. Acad. Sci. U.S.A. 82, 6379-6383, 1985
A:Title: Structure of canine pulmonary surfactant apoprotein: cDNA and complete amino
A:Reference number: A5296; MUID:86016705
A:Accession: A5296
A:Molecule type: mRNA
A:Residues: 1-248 <BEN>
A:Note: the authors translated the codon TGC for residue 60 as Pro
A:Note: part of the sequence, including the amino end of the mature protein, was confirmed
R:Liau, D.F.; Ryan, S.F.
Chem. Phys. Lipids 59, 29-38, 1991
A:Title: Purification of surfactant protein A from dog lung by reconstitution with su
A:Reference number: A61227; MUID:92163993
A:Accession: A61227
A:Molecule type: protein
A:Residues: 18-32 <LIA>
R:Ross, G.F.; Meuth, J.; Ohning, B.; Kim, Y.; Whitsett, J.A.
Biochim. Biophys. Acta 870, 267-278, 1986
A:Title: Purification of canine surfactant-associated glycoproteins A. Identification
A:Reference number: A60142; MUID:86159848
A:Accession: A60142
A:Molecule type: protein
A:Residues: 24-34:95-101, X', 103-108 <ROS>
R:Patthy, L.
Nature 325, 490, 1987
A:Reference number: A93388; MUID:87115834
A:Contents: annotation; animal lectin domain homology
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower
pendent on the presence of calcium ions
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: alveolar proteinosis; calcium; disulfide bond; gaseous exchange; hydroxypro
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-248/Product: pulmonary surfactant protein A #status experimental <MPT>
F:127-246/Domain: collagen-like
F:20,207/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:30/Modified site: 4-hydroxyproline (Pro) #status experimental

Est

A:Cross-references: GB:M61231; NID:g207035; PIDN:AAA42170.1; PID:g207036
 A:Experimental source: lung
 A>Note: sequence extracted from NCBI backbone (NCBI:76027, NCBI:P:76031)
 C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
 F:253-372/Domain: C-type lectin homology <LCH>

Query Match 66.5%; Score 103; DB 1; Length 374;
 Best Local Similarity 59.3%; Pred. No. 2.5e-07;
 Matches 16; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCLOSLRAICEF 27
 |||||:||||:||||:||||:||||
 Db 348 ENCVEIFPDGKNWDPKSLQILLICEF 374

RESULT 12

S33603
 surfactant protein D - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
 C:Accession: S33603
 R:Lim, B.L.; Lu, J.; Reid, K.B.M.
 Immunology 78, 159-165, 1993
 A>Title: Structural similarity between bovine conglutinin and bovine lung surfactant protein D
 A:Reference number: S33603; MUID:93170856
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-369 <LIM>
 C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
 F:248-367/Domain: C-type lectin homology <LCH>

Query Match 64.5%; Score 100; DB 2; Length 369;
 Best Local Similarity 59.3%; Pred. No. 6.9e-07;
 Matches 16; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCLOSLRAICEF 27
 |||||:||||:||||:||||:||||
 Db 343 ENCVEIFPDGKNWDPKSLQILLICEF 369

RESULT 13

JN0450
 conglutinin precursor - bovine
 N:Alternate names: C3b-binding protein
 N:Contains: conglutinin-N
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 R:Suzuki, Y.; Yin, Y.P.; Makino, M.; Kurimura, T.; Wakamiya, N.
 Biochem. Biophys. Res. Commun. 191, 335-342, 1993
 A>Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.
 A:Reference number: JN0450; MUID:93213261
 A:Accession: JN0450
 A:Molecule type: mRNA
 A:Residues: 1-371 <SUZ>
 A:Cross-references: DDBJ:D14085; NID:g285643; PIDN:BA03170.1; PID:g285644
 A:Experimental source: liver
 R:Kawasaki, N.; Itoh, N.; Kawasaki, T.
 Biochem. Biophys. Res. Commun. 198, 597-604, 1994
 A>Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mammalian gene
 A:Reference number: JC2396; MUID:94128104
 A:Accession: JC2396
 A:Molecule type: mRNA
 A:Residues: 1-371 <KA2>
 A>Note: The authors translated the codon GAT for residues 250 and 270 as Glu
 R:Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.
 Biochem. J. 292, 157-162, 1993
 A>Title: The cDNA cloning of conglutinin and identification of liver as a primary site of synthesis
 A:Reference number: S33235; MUID:93277452
 A:Accession: S33235

A:Molecule type: mRNA
 A:Residues: 1-172 'H'; 174-217, 'A'; 219-271, 'V'; 273-371 <LUJ>
 A:Cross-references: EMBL:X71774; NID:g395267; PIDN:CAA50665.1; PID:g395268
 A:Experimental source: liver
 R:Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.
 J. Biol. Chem. 266, 2715-2723, 1991
 A>Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin family

A:Reference number: A23740; MUID:91131556
 A:Accession: A23740
 A:Molecule type: protein
 A:Residues: 21-209, 'S'; 211-371 <LEE>
 R:Kawasaki, N.; Tokota, Y.; Kawasaki, T.
 Arch. Biochem. Biophys. 305, 533-540, 1993
 A>Title: Differentiation of conglutination activity and sugar-binding activity of conglutinin
 A:Reference number: S36879; MUID:93384312
 A:Accession: S36879

A:Molecule type: protein
 A:Residues: 21-54; 75-86, 'X'; 88-89, 'X'; 91, 'X'; 93-94; 208-209, 'X'; 211-227 <KAW>
 A:Experimental source: serum
 R:Lu, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.
 Eur. J. Biochem. 215, 793-799, 1993
 A>Title: Structural similarity between lung surfactant protein D and conglutinin. Two different proteins
 A:Reference number: S35044; MUID:93358905
 A:Accession: S35044

A:Molecule type: protein
 A:Residues: 75-86, 'X'; 88-89, 'X'; 91, 'I' <LUA>
 A:Experimental source: lung
 R:Young, N.M.; Leon, M.A.
 Biochem. Biophys. Res. Commun. 143, 645-651, 1987
 A>Title: The carbohydrate specificity of conglutinin and its homology to proteins in the conglutinin family
 A:Reference number: A29416; MUID:87184551
 A:Contents: annotation
 R:Malhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.
 Biochem. J. 293, 15-19, 1993

A>Title: Research Communication. Localization of the receptor-binding site in the conglutinin family
 A:Reference number: S34054; MUID:93319501
 A:Contents: annotation
 R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sas
 J. Immunol. 153, 173-180, 1994
 A>Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship with other conglutinins
 A:Reference number: I46010; MUID:94267222
 A:Accession: I46010

A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-371 <LIO>
 A:Cross-references: EMBL:U06860; NID:g507183; PIDN:AAB60624.1; PID:g514256
 C:Comment: This protein mediates the agglutination of erythrocytes with antibody and is a Ca²⁺-dependent serum lectin specific for N-acetylglucosamine
 C:Genetics:

A:Gene: CGN1
 A:Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1
 C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
 C:Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-371/Product: conglutinin #status predicted <MAT>
 F:46-214/Region: collagen-like
 F:75-371/Product: conglutinin-N #status predicted <KA2>
 F:248-369/Domain: C-type lectin homology <LCH>
 F:63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #status
 F:63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxylysine (Lys) #status exper
 F:78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (Pro) #status
 F:337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 61.3%; Score 95; DB 1; Length 371;
 Best Local Similarity 59.3%; Pred. No. 3.7e-06;
 Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCLOSLRAICEF 27
 |||||:||||:||||:||||:||||
 Db 345 ENCVEIFPDGKNWDPKSLQILLICEF 371

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:40:43 ; Search time 15.79 Seconds
(without alignments)
66.208 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVMYTDGKWNDRCLQSLRAICEF 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	142	91.6	247	1 PSPA_RABIT	P12842 oryctolagus
2	140	90.3	248	1 PSPA_MOUSE	P35242 mus musculus
3	139	89.7	249	1 PSPA_PIG	P49874 sus scrofa
4	137	88.4	248	1 PSPA_HUMAN	P07714 homo sapien
5	136	87.7	248	1 PSPA_RAT	P08427 rattus norv
6	133	85.8	248	1 PSPA_CANFA	P06908 canis fami
7	132	85.2	247	1 PSPA_CAYPO	P50403 cavia porce
8	108	69.7	375	1 PSPD_HUMAN	P35247 homo sapien
9	103	66.5	301	1 CL43_BOVIN	P42916 bos taurus
10	103	66.5	374	1 PSPD_MOUSE	P50404 mus musculus
11	103	66.5	374	1 PSPD_RAT	P35248 rattus norv
12	100	64.5	369	1 PSPD_BOVIN	P35246 bos taurus
13	95	61.3	371	1 CONG_BOVIN	P23805 bos taurus
14	87	56.1	244	1 MABC_RAT	P08661 rattus norv
15	83	53.5	244	1 MABC_MOUSE	P41117 mus musculus
16	81	52.3	248	1 MABC_HUMAN	P11226 homo sapien
17	81	52.3	249	1 MABC_BOVIN	P02659 bos taurus
18	80.5	51.9	485	1 LEM2_BOVIN	P98107 bos taurus
19	79.5	51.3	484	1 LEM2_PIG	P98110 sus scrofa
20	79.5	51.3	830	1 LEM3_HUMAN	P16109 homo sapien
21	79	51.0	311	1 LEC1_HUMAN	P07307 homo sapien
22	77.5	50.0	611	1 LEM2_CANFA	P33730 canis fami
23	77	49.7	238	1 MABA_RAT	P19999 rattus norv
24	72.5	46.8	551	1 LEM2_RABIT	P27113 oryctolagus
25	72.5	46.8	610	1 LEM2_HUMAN	P16581 homo sapien
26	72.5	46.8	612	1 LEM2_MOUSE	P00690 mus musculus
27	72	46.5	301	1 LEC1_MOUSE	P24721 mus musculus
28	71.5	46.1	549	1 LEM2_RAT	P98105 rattus norv
29	71	45.8	239	1 MABA_MOUSE	P39039 mus musculus
30	70.5	45.5	370	1 LEM1_BOVIN	P98131 bos taurus
31	69.5	44.8	768	1 LEM3_MOUSE	P01102 mus musculus
32	69.5	44.8	768	1 LEM3_RAT	P98106 rattus norv
33	68.5	44.2	372	1 LEM1_MOUSE	P18337 mus musculus

34	68.5	44.2	372	1 LEM1_RAT	P30836 rattus norv
35	67.5	43.5	372	1 LEM1_MACMU	Q95198 macaca mulia
36	67.5	43.5	372	1 LEM1_PAPHA	Q28768 papio hamad
37	67.5	43.5	646	1 LEM3_BOVIN	P42201 bos taurus
38	67	43.2	301	1 LEC1_RAT	P08290 rattus norv
39	66.5	42.9	372	1 LEM1_HUMAN	P14151 homo sapien
40	66.5	42.9	372	1 LEM1_PANTR	Q95237 pan troglod
41	66.5	42.9	372	1 LEM1_PONPY	Q95235 pongo pygma
42	65.5	42.3	769	1 LEM3_SHEEP	P98109 ovis aries
43	65	41.9	147	1 PL1A_TRIFL	P21755 trimeresuru
44	65	41.9	197	1 CLF1_HUMAN	Q75596 homo sapien
45	65	41.9	283	1 LEC1_MOUSE	P34927 mus musculus

ALIGNMENTS

```

RESULT 1
PSPA_RABIT
ID PSPA_RABIT STANDARD; PRT; 247 AA.
AC P12842;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP).
GN SFTPAL OR SFTPA OR SFTPL
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88139346; PubMed=2830270;
RA Boggaram V., Qing K., Mendelson C.R.;
RT "The major apoprotein of rabbit pulmonary surfactant. Elucidation of
RT primary sequence and cyclic AMP and developmental regulation.";
RL J. Biol. Chem. 263:2939-2947(1988).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=92312742; PubMed=1616051;
RA Chen Q., Boggaram V., Mendelson C.R.;
RT "Rabbit lung surfactant protein A gene: identification of a lung-
RT specific DNase I hypersensitive site.";
RL Am. J. Physiol. 262:L682-L671(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03542; AAA31465.1; -
CC EMBL; L19187; AAA31468.1; -
CC EIR; A29931; LNRBPS.
CC HSP; P35247; I808.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR001304; lectin_c.

```



```
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS00041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT DOMAIN 27 99 A.
FT DOMAIN 152 247 COLLAGEN-LIKE.
FT DISULFID 154 245 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 223 237 BY SIMILARITY.
FT CARBOHYD 206 206 BY SIMILARITY.
FT VARIANT 12 12 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 57 60 S -> P.
FT SEQUENCE 247 AA; 25071 MW; 289634054CBC8CBA CRC64;
Query Match 91.6%; Score 142; DB 1; Length 247;
Best Local Similarity 88.9%; Pred. No. 1.le-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 EKCVMYTDGKNDKNCLOSLRAICEF 27
|||||
DB 221 EKCVMYTDGKNDKNCLOSLRAICEF 247
|||||
RESULT 2
PSPA_MOUSE STANDARD; PRT; 248 AA.
AC P35242:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP).
GN SFTPA1 OR SFTPA OR SFTP1 OR SFTP-1.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DNA/2J.
RX MEDLINE=93072386; PubMed=1443158;
RA Korfhagen T.R., Bruno M.D., Glasser S.W., Ciraoilo P.J., Whitsett J.A.,
RA Lattier D.L., Wikenheiser K.A., Clark J.C.;
RT "Murine pulmonary surfactant SP-A gene: cloning, sequence, and
RT transcriptional activity.";
RL Am. J. Physiol. 263:L546-L554(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S48766; AAB24274.1; -.
```

```
DR PIR: A48853; A48853.
DR HSP: P35247; IB08.
DR MGD; MGI:109518; Sftpa.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS00041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT DOMAIN 28 100 COLLAGEN-LIKE.
FT DOMAIN 153 248 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 155 246 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (PROBABLE).
FT SEQUENCE 248 AA; 26157 MW; 5688BF070E3EB9AE CRC64;
Query Match 90.3%; Score 140; DB 1; Length 248;
Best Local Similarity 88.9%; Pred. No. 2.le-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 EKCVMYTDGKNDKNCLOSLRAICEF 27
|||||
DB 222 EKCVMYTDGKNDKNCLOSLRAICEF 248
|||||
RESULT 3
PSPA_PIG STANDARD; PRT; 249 AA.
AC P49874:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP).
GN SFTPA1 OR SFTPA OR SFTP1.
OS Sus scrofa (Pig).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Adamou J.E., Elshourbagy N.A.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C). (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L41350; AAA88403.1; -.
```

```
DR HSP: P19999; LYTT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE: PS00411; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 249
FT DOMAIN 28 100
FT DOMAIN 153 249
FT DISULFID 155 247
FT DISULFID 225 239
FT CARBOHYD 208 208
SQ SEQUENCE 249 AA; 26702 MW; 3C4E05AD07F2A7CD CRC64;

Query Match 89.7%; Score 139; DB 1; Length 249;
Best Local Similarity 88.9%; Pred. No. 3e-13; Indels 0; Gaps 0;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDNCQLSLAICEF 27
      |||||:|||||:|||||
DB 223 EKCVMYTDGQWDRNCQYRLAICEF 249

RESULT 4
ID PSPA_HUMAN STANDARD; PRT; 248 AA.
AC P07714;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSP) (Alveolar proteinosis protein) (35 kDa pulmonary surfactant-
DE associated protein).
GN SFTPA1 OR SFTPA OR SFTPL1 OR PSPAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86250832; PubMed=3755136;
RA Floros J., Steinbrink R., Jacobs K., Phelps D., Kriz R., Recny M.,
RA Sultman L., Jones S., Taesch H.W., Frank H.A., Fritsch E.F.;
RT "Isolation and characterization of cDNA clones for the 35-kDa
RT pulmonary surfactant-associated protein.";
RL J. Biol. Chem. 261:9029-9033(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014366; PubMed=2995821;
RA White R.T., Damm D., Miller J., Spratt K., Schilling J., Hawgood S.,
RA Benson B., Cordell B.;
RT "Isolation and characterization of the human pulmonary surfactant
RT apoprotein gene.";
RL Nature 317:361-363(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92198680; PubMed=1372511;
RA Katyal S.L., Singh G., Locker J.L.;
RT "Characterization of a second human pulmonary surfactant-associated
RT protein SP-A gene.";
RL Am. J. Respir. Cell Mol. Biol. 6:446-452(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
```

```
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13686; AAA60211.1; -
CC EMBL; K03475; AAA36520.1; -
CC EMBL; M30838; AAA36510.1; -
CC EMBL; M68519; AAA60319.1; -
CC PIR; A24622; LNHUP5.
CC PIR; A25720; LNHUP6.
CC PIR; B25720; LNHUP1.
CC HSP; P19999; LYTT.
CC MIM; 178630; -
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001304; lectin_c.
CC Pfam; PF00059; lectin_c; 1.
CC SMART; SM00034; CLECT; 1.
CC PROSITE; PS00615; C-TYPE LECTIN_1; 1.
CC PROSITE; PS00411; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat; Polymorphism.
FT SIGNAL 1 20
FT CHAIN 21 248
FT DOMAIN 28 100
FT DOMAIN 153 248
FT DISULFID 155 246
FT DISULFID 224 238
FT CARBOHYD 207 207
FT VARIANT 9 9
FT VARIANT 50 50
FT VARIANT 66 66
FT VARIANT 73 73
FT VARIANT 81 81
FT VARIANT 85 85
FT VARIANT 219 219
FT VARIANT 223 223
FT VARIANT 19 19
FT VARIANT 45 45
FT VARIANT 54 54
FT VARIANT 91 91
FT VARIANT 100 100
FT VARIANT 247 247
FT SEQUENCE 248 AA; 26214 MW; 6A9F0C3498BF3633 CRC64;

Query Match 88.4%; Score 137; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 5.8e-13;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDNCQLSLAICEF 27
      |||||:|||||:|||||
DB 222 EKCVMYTDGQWDRNCQLSLAICEF 248
```

DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
 KW Signal; Lectin; Collagen; Repeat.
 KWK SIGNAL
 KWK SIGNAL 1 20
 CHAIN 21 248
 PULMONARY SURFACTANT-ASSOCIATED PROTEIN A.

	DOMAIN	28	100
FFTF	DOMA1N	153	248
FFTF	D1SULFID	155	246
FFTF	D1SULFID	224	238
FFTF	CARBOHYD	21	21
FFTF	CARBOHYD	207	207
FFTF	CONFLICT	78	78
FFTF	CONFLICT	84	84
FFTF	CONFLICT	139	139
FFTF	CONFLICT	156	156
FFTF	CONFLICT	180	180
SSQ	SEQUENCE	248 AA;	26288 MW;

Query Match 87.7%; Score 136; DB 1; Length 248;
 Best Local Similarity 85.2%; Pred.No. 8,1e-13;
 Matches 23; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 EKCVMYTDGKWNDRNCQLSRLAICEF 27
 |||||
 Db 222 EKCVMYTDGKWNDRGCLOYRLAVCEF 248

RESULT 6
 PSPA_CANFA STANDARD; PRT: 248 AA.
 AC P06908;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A) (PSAP).
 DE SFTPA1 OR SFTPA OR SFTPL.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_Taxid=9615;
 [1]
 RA SEQUENCE FROM N.A.; AND SIGNAL SEQUENCE CLEAVAGE SITE.
 RP MEDLINE=86016705; PubMed=3863100;
 RR Benson B., Hawgood S., Schilling J., Clements J., Damm D., Cordell B.,
 RA White R.T.:
 RA "Structure of canine pulmonary surfactant apoprotein: cDNA and
 RT complete amino acid sequence.";
 RT Proc. Natl. Acad. Sci. U.S.A. 82:6379-6383(1985).
 [2]
 RP C-TYPE LECTIN DOMAIN.
 RP MEDLINE=87115834; PubMed=3808053;
 RA Patti L.;
 RA "Is lung surfactant protein a lectin-collagen hybrid?";
 RP Nature 325:490-490(1987).
 RL -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
 CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
 CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
 CC ESSENTIAL FOR NORMAL RESPIRATION.
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation ?

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M11769; AAB30887.1; -
DR PIR: A25296; LNDGSP.
DR HSP: P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; LECTIN_C.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00059; LECTIN_C; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; LECTIN; Collagen; Repeat.
FT SIGNAL 1 17
FT CHAIN 18 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT A.
FT DOMAIN 28 100 COLLAGEN-LIKE.
FT DOMAIN 153 248 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 155 246 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (PROBABLE).
SQ SEQUENCE 248 AA; 26268 MW; 340FE95D4E2502C0 CRC64;

Query Match 85.8%; Score 133; DB 1; Length 248;
Best Local Similarity 81.5%; Pred. No. 2.2e-12;
Matches 22; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
|:|||||:|:|||||:|:|||||
Db 222 EQCVEMTDGQNNKNCLOSLRAICEF 248

RESULT 7

ID PSPA_CAVPO STANDARD; PRT: 247 AA.
AC P50403;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSP-A).
GN SFTPA1 OR SFTPA OR SFTPL.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY; TISSUE=Lung;
RX MEDLINE=98018900; PubMed=9357868;
RA Xuan H.T., Gowan S., Kelly F.J., Bingle C.D.;
RT "Cloning of guinea pig surfactant protein A defines a distinct
RT cellular distribution pattern within the lung.";
RL Am. J. Physiol. 273:1900-1906(1997).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, SPAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U40869; AAB82952.1; -
DR HSP: P19993; 1YTT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; LECTIN_C.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00059; LECTIN_C; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; LECTIN; Collagen; Repeat.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 247 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT A.
FT DOMAIN 27 99 COLLAGEN-LIKE.
FT DOMAIN 152 245 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 154 245 BY SIMILARITY.
FT DISULFID 223 237 BY SIMILARITY.
FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 247 AA; 26104 MW; D1BC86270EEFC932 CRC64;

Query Match 85.2%; Score 132; DB 1; Length 247;
Best Local Similarity 81.5%; Pred. No. 3.1e-12;
Matches 22; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
|:|||||:|:|||||:|:|||||
Db 221 EKCAEMYLDGTWNDKNCLOSLRAICEF 247

RESULT 8

ID PSPD_HUMAN STANDARD; PRT: 375 AA.
AC P35247;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPA OR PSPD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93155122; PubMed=8428971;
RA Crouch E., Rust K., Veille R., Donis-Keller H., Grosso L.;
RT "Genomic organization of human surfactant protein D (SP-D). SP-D is
RT encoded on chromosome 10q22.2-23.1.";
RL J. Biol. Chem. 268:2976-2983(1993).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 214-243.
RC TISSUE=Lung, and Amniotic fluid;
RX MEDLINE=92322003; PubMed=1339284;
RA Lu J., Wallis A.C., Reid K.B.M.;
RT "Purification, characterization and cDNA cloning of human lung
RT surfactant protein D.";
RL Biochem. J. 284:795-802(1992).
RN [3]
RP SEQUENCE OF 60-375 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Lung;

```

SQ SEQUENCE 375 AA; 37702 MW; 2986B2699FC01A6A CRC64;

Query Match 59.7%; Score 108; DB 1; Length 375;
Best Local Similarity 63.0%; Pred. No. 1.5e-08;
Matches 17; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKVENVYTDGKWNDRNCIASRLAICEF 27
| | | | : | | | | | | | | : | | | |
Db 349 EDCEVEITNGKWDRCAGEKRUVCEEF 375

RESULT 9
CL43_BOVIN STANDARD; PRT; 301 AA.
AC P42916;
AD 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Collectin-43 (CL-43).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Liver;
RX MEDLINE=94216283; PubMed=8163480;
RA Lim B.-L., Willis A.C., Reid K.B.M., Lu J., Laursen S.B.,
RA Jensenius J.C., Holmskov U.;
RT "Primary structure of bovine collectin-43 (CL-43). Comparison with
RT conglutinin and lung surfactant protein-D.";
RL J. Biol. Chem. 269:11820-11824(1994).
CC -!- FUNCTION: LECTIN THAT BINDS TO VARIOUS SUGARS: MANNOSE >
CC FUCOSE > GLCNAC > GLUCOSE = GALACTOSE > LACTULOSE >
CC GALANAL. COULD PLAY A ROLE IN IMMUNE DEFENSE.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DB EMBL; X75912; CAA53511.1; ALT_SEQ.
DR HSSP; P35247; 1B08.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; Lectin_c.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SW00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Collagen;
KW Repeat; Calcium.
FT DOMAIN 29 142 COLLAGEN-LIKE.
FT FT 301 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 204 299 BY SIMILARITY.
FT DISULFID 277 291 BY SIMILARITY.
FT SEQUENCE 301 AA; 31362 MW; 0385C10B9424CD76 CRC64;

Query Match 66.5%; Score 103; DB 1; Length 301;
Best Local Similarity 59.3%; Pred. No. 6.6e-08;
Matches 16; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 EKVENVYTDGKWNDRNCIASRLAICEF 27

```

Qy	1	EKCVENYTDGKWNDRNCLQSR	LAICEF	27
		: : :	:	
Db	348	ENCVEIFTNGOWNDKACGEOR	LVICEF	374

RESULT	ID	PROTEIN	STANDARD	PRT	374 AA
AC	P35248	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DE	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D) (CP4).				
GN	SFTPD OR SFTP4				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RC	SEQUENCE FROM N.A., AND SEQUENCE OF 20-33.				
RP	TISSUE=Lung;				
RX	MEDLINE=92112913; PubMed=1370483;				
RA	Shimizu H., Fisher J.H., Papst P., Benson B., Lau K., Mason R.J.,				
RA	Voelkel D.R.;				
RT	"Primary structure of rat pulmonary surfactant protein D. cDNA and				
RT	deduced amino acid sequence."				
RL	J. Biol. Chem. 267:1853-1857(1992).				
RL	[2]				
RP	SEQUENCE OF 73-95 AND 153-180.				
RC	TISSUE=Lung;				
RX	MEDLINE=90001186; PubMed=2675969;				
RA	Persson A., Chang D., Rust K., Moxley M., Longmore W., Crouch E.;				
RA	"Purification and biochemical characterization of CP4 (SP-D), a				
RL	collagenous surfactant-associated protein."				
RL	Biochemistry 28:6361-6367(1989).				
CC	!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED				
CC	MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER				
CC	EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE				
CC	EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.				
CC	!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.				
CC	!- SUBCELLULAR LOCATION: Extracellular.				
CC	!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%				
CC	PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,				
CC	CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL				
CC	HYDROPHOBIC PROTEINS (SP-B AND SP-C).				
CC	!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.				
CC	!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/)				
CC	or send an email to license@sib-sib.ch .				
CC	-----				
DR	EMBL; M81231; AAA42170.1; -				
DR	PIR; A42046; A42046.				
DR	HSSP; P35247; 1B08.				
DR	InterPro; IPR000087; Collagen.				
DR	InterPro; IPR001304; lectin_c.				
DR	Pfam; PF01391; Collagen; 3.				
DR	Pfam; PF00059; lectin_c; 1.				
DR	SMART; SM00034; CLECT; 1.				
DR	PROSITE; PS00615; C-TYPE LECTIN.1; 1.				
DR	PROSITE; PS00041; C-TYPE LECTIN.2; 1.				
KW	Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;				
KW	Signal; Lectin; Collagen; Repeat; Coiled coil.				
FT	SIGNAL 1 19				
FT	CHAIN 20 374 PULMONARY SURFACTANT-ASSOCIATED PROTEIN				
FT	FT D.				

```
Query Match      66.5%; Score 103; DB 1; Length 374;
Best Local Similarity 59.3%; Pred. No. 8.3e-08;
Matches 16; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
```

```

ET DOMAIN 45 221 COLLAGEN-LIKE.
ET DOMAIN 222 253 COILED COIL (POTENTIAL).
ET DOMAIN 278 374 C-TYPE LECTIN (SHORT FORM).
ET DISULFID 280 372 BY SIMILARITY.
FT DISULFID 350 364 BY SIMILARITY.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .).
FT MOD_RES 77 77 HYDROXYLATION.
FT MOD_RES 86 86 HYDROXYLATION.
FT MOD_RES 95 95 HYDROXYLATION.
FT MOD_RES 98 98 HYDROXYLATION.
FT MOD_RES 170 170 HYDROXYLATION.
FT MOD_RES 176 176 HYDROXYLATION.
FT CONFLICT 89 89 N -> E (IN REF. 2).
FT CONFLICT 164 164 K -> C (IN REF. 2).
SQ SEQUENCE 374 AA; DB2BB5E399DB4A3C CRC64;

Query Match 66.5%; Score 103; DB 1; Length 374;
Best Local Similarity 59.3%; Pred. No. 8.3e-08;
Matches 16; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKVEMVTDGKWNDRNCLOSRLAICEF 27
| | | | | : | | | | | : | | | | |
Db 348 ENCVEIFPGKWNDRKACGEORLVICEF 374

RESULT 12
PSPD_BOVIN
ID PSPD_BOVIN STANDARD; PRT; 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
RC TISSUE=Lung;
RX MEDLINE=93170856; PubMed=8436402;
RA Lim B.L., Lu J., Reid K.B.M.;
RT "Structural similarity between bovine conglutinin and bovine lung surfactant protein D and demonstration of liver as a site of synthesis of conglutinin.";
RT Immunology 78:159-165(1993).
CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75911; CAA53510.1;
CC PIR; S33603; S33603.
CC HSSP; P35247; 1B08.

```

```

DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00641; C-TYPE LECTIN_2; 1.
RW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat; Coiled coil.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 369 PULMONARY SURFACTANT-ASSOCIATED PROTEIN D.
ET DOMAIN 46 216 COLLAGEN-LIKE.
ET DOMAIN 217 248 COILED COIL (POTENTIAL).
ET DOMAIN 273 369 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 275 367 BY SIMILARITY.
FT DISULFID 345 359 BY SIMILARITY.
ET CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
ET MOD_RES 78 78 HYDROXYLATION (BY SIMILARITY).
ET MOD_RES 87 87 HYDROXYLATION (BY SIMILARITY).
ET MOD_RES 96 96 HYDROXYLATION (BY SIMILARITY).
ET MOD_RES 99 99 HYDROXYLATION (BY SIMILARITY).
ET MOD_RES 165 165 HYDROXYLATION (BY SIMILARITY).
ET MOD_RES 171 171 HYDROXYLATION (BY SIMILARITY).
SQ SEQUENCE 369 AA; 37361 MW; 07D88B24E0AEB2E3 CRC64;

Query Match 64.5%; Score 100; DB 1; Length 369;
Best Local Similarity 59.3%; Pred. No. 2.2e-07;
Matches 16; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 EKVEMVTDGKWNDRNCLOSRLAICEF 27
| | | | | : | | | | | : | | | | |
Db 343 ENCVEIFPGKWNDRKACGEORLVICEF 369

RESULT 13
CONG_BOVIN
ID CONG_BOVIN STANDARD; PRT; 371 AA.
AC P23805;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Conglutinin precursor.
GN CGNI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=93213261; PubMed=8460993;
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RP MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;
RT "The cDNA cloning of conglutinin and identification of liver as a primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
[3]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=94215917; PubMed=8163202;
RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okama T.B., Tauber A.I., Sastry K.N.;
RT "Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of the BC cDNA reveals strong homology to surfactant protein-D.";
RL Gene 141:277-281(1994).

```

FT	MOD_RES	198	198	HYDROXYLATION.
FT	SITE	201	203	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	275	369	BY SIMILARITY.
FT	DISULFID	347	361	BY SIMILARITY.
FT	CARBOHYD	337	337	N-LINKED (GLCNAC. -) (POTENTIAL).
FT	CONFLICT	173	173	R -> H (IN REF. 2 AND 3).
FT	CONFLICT	210	210	K -> S (IN REF. 5).
FT	CONFLICT	218	218	V -> A (IN REF. 2).
FT	CONFLICT	272	272	E -> V (IN REF. 2).
FT	SEQUENCE	371 AA:	37994 MW;	867BB41992544BIF CRC64;
Query Match 61.3%; Score 95; DB 1; Length 371;				
Best Local Similarity 59.3%; Pred.No. 1.2e-06;				
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps				
QY	1 EKVEMYTGKWNDRNCQLSRALICEF 27			
	I L I I I I I I I I I I I : I I I I I			
Db	345 ENCVEIFPDGKWNDVPCSKQLLVICFE 371			
RESULT 14				
ID	MABC_RAT	STANDARD;	PRT;	244 AA.
AC	P08661;			
DT	01-JAN-1988 (Rel. 05, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DE	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)			
GN	(RA-reactive factor P28A subunit) (RARE/P28A).			
DN	MBL2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
[1]				
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=92296655; PubMed=1607365;			
RX	Wada M., Itoh N., Ohta M., Kawasaki T.;			
RA	"Characterization of rat liver mannan-binding protein gene.";			
RL	J. Biochem. 111:66-73(1992).			
[2]				
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=86196130; PubMed=3009480;			
RA	Drickamer K., Dordal M.S., Reynolds L.;			
RA	"Mannose-binding proteins isolated from rat liver contain			
RT	carbohydrate-recognition domains linked to collagenous tails.			
RT	complete primary structures and homology with pulmonary surfactant			
RL	apoprotein.";			
RL	J. Biol. Chem. 261:6878-6887(1986).			
[3]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Wistar; TISSUE=Liver;			
RX	MEDLINE=87194686; PubMed=3032924;			
RA	Oka S., Itoh N., Kawasaki T., Yamashina I.;			
RA	"Primary structure of rat liver mannan-binding protein deduced from			
RT	its cDNA sequence.";			
RL	J. Biochem. 101:135-144(1987).			
[4]				
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 133-244.			
RN	MEDLINE=96132792; PubMed=8557671;			
RA	Ng K.K.-S., Drickamer K., Weis W.I.;			
RA	"Structural analysis of monosaccharide recognition by rat liver			
RT	mannose-binding protein.";			
RL	J. Biol. Chem. 271:663-674(1996).			
CC	!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-			
CC	DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,			
CC	BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF			
CC	THE ANTIBODY.			
CC	!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.			
CC	!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR. MOST			

CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
 CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
 CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNA
 CC SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE
 CC PROTEIN.
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M14103; AAA41554.1; -.
 CC EMBL; X05023; CAA28687.1; -.
 CC PIR; A24791; LNRPMC.
 CC PIR; A26798; A26798.
 CC PIR; JX0201; JX0201.
 CC PDB; 1RDI; 08-MAR-96.
 CC PDB; 1RDJ; 08-MAR-96.
 CC PDB; 1ROK; 08-MAR-96.
 CC PDB; 1ROJ; 08-MAR-96.
 CC PDB; 1ROK; 08-MAR-96.
 CC PDB; 1ROJ; 08-MAR-96.
 CC PDB; 1ROK; 08-MAR-96.
 CC PDB; 1ROJ; 08-MAR-96.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001304; lectin_c.
 CC Pfam; PF000059; lectin_c; 1.
 CC SMART; SM00034; CLECT; 1.
 CC PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 CC PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 CC Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;
 KW Calcium; Collagen; Repeat; Signal; 3D-structure.
 FT SIGNAL 1 18
 FT CHAIN 19 244 MANNOSE-BINDING PROTEIN C.
 FT DOMAIN 38 96 COLLAGEN-LIKE.
 FT DOMAIN 149 242 C-TYPE LECTIN (SHORT FORM).
 FT MOD_RES 43 43 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 58 58 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 69 69 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 78 78 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 81 81 HYDROXYLATION (POTENTIAL).
 FT DISULFID 29 29 INTERCHAIN.
 FT DISULFID 34 34 INTERCHAIN.
 FT DISULFID 151 240
 FT DISULFID 218 232
 FT CONFLICT 38 39 GL -> AW (IN REF. 2).
 SQ SEQUENCE 244 AA; 26014 MW; F0706E2AA9331531 CRC64;
 Query Match 56.1%; Score 87; DB 1; Length 244;
 Best Local Similarity 55.6%; Pred. NO. 1.2e-05;
 Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
 QY 1 EKCVEMVTDCKWNRNCIQSLAICEF 27
 DB 216 ENCVLLTGKNDVPCDSFLVWCFE 242
 RESULT 15
 ID MABC_MOUSE STANDARD; PRT; 244 AA.
 AC P4117;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)
 DE (RA-reactive factor P28A subunit) (RARE/P28A).
 GN MBP.
 OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBA/J; TISSUE=Liver;
 RX MEDLINE=91302823; PubMed=1712818;
 RA Sastry K., Zahedi K., Lellias J.M., Whitehead A.S., Ezekowitz R.A.;
 RT "Molecular characterization of the mouse mannose-binding proteins.
 RL The mannose-binding protein A but not C is an acute phase reactant.";
 RJ J. Immunol. 147:692-697(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C;
 RX MEDLINE=95284466; PubMed=7766991;
 RA Sastry R., Wang J.S., Brown D.C., Ezekowitz R.A., Tauber A.I.,
 RA Sastry K.N.;
 RT "Characterization of murine mannose-binding protein genes Mb11 and
 RL Mb12 reveals features common to other collectin genes.";
 RJ Mamm. Genome 6:103-110(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kuge S., Ihara S., Watanabe E., Watanabe M., Takishima K., Suga T.,
 RA Mamaiya G., Kawakami M.;
 RA Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-
 CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,
 CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF
 CC THE ANTIBODY.
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
 CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
 CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
 CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
 CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNA
 CC SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE
 CC PROTEIN.
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; S42294; AAB19343.1; -.
 CC EMBL; U09016; AAA82010.1; -.
 CC EMBL; U09013; AAA82010.1; JOINED.
 CC EMBL; U09014; AAA82010.1; JOINED.
 CC EMBL; U09015; AAA82010.1; JOINED.
 CC EMBL; D11440; BAA02005.1; -.
 CC HSSP; P08661; IPDO.
 CC MGD; MGI:96924; Mb12.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001304; lectin_c.
 CC Pfam; PF000059; lectin_c; 1.
 CC SMART; SM00034; CLECT; 1.
 CC PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 CC PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 CC Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;
 KW Calcium; Collagen; Repeat; Signal.
 FT SIGNAL 1 18 BY SIMILARITY.
 FT CHAIN 19 244 MANNOSE-BINDING PROTEIN C.
 FT DOMAIN 38 96 COLLAGEN-LIKE.
 FT DOMAIN 149 242 C-TYPE LECTIN (SHORT FORM).
 FT MOD_RES 43 43 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 58 58 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 69 69 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 78 78 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 81 81 HYDROXYLATION (POTENTIAL).
 FT DISULFID 29 29 INTERCHAIN (BY SIMILARITY).

FT DISULFID 34 34 INTERCHAIN (BY SIMILARITY).
FT DISULFID 151 240 BY SIMILARITY.
FT DISULFID 218 232 BY SIMILARITY.
FT CONFLICT 3 3 I -> I (IN REF. 1).
FT CONFLICT 15 15 V -> A (IN REF. 1).
SQ SEQUENCE 244 AA; 25957 MW; 49AE84E2290DEB0A CRC64;

Query Match 53.5%; Score 83; DB 1; Length 244;
Best Local Similarity 59.3%; Pred No. 4.4e-05;
Matches 16; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 EKVEMYTDGKWNDRNCLQSLAICEF 27
| | | : | | | | | | | | | |
Db 216 EDCVVILGNGKWNVDVPCSDSFLAICEF 242

Search completed: July 3, 2002, 12:40:43
Job time: 351 sec

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1	352.5	23.8	369	2	A33603	surfactant protein
2	2	345	23.2	301	2	A53570	collectin-43 - bov
3	3	337	22.7	375	1	A45225	pulmonary surfacta
4	4	333	22.4	374	1	A42046	surfactant protein
5	5	328.5	22.1	371	1	JN0450	conglutinin precu
6	6	328.5	22.1	371	2	I45878	conglutinin - bov1
7	7	305.5	20.6	247	1	LNRGPs	pulmonary surfacta
8	8	290	19.5	238	1	LNRMC	mannose-binding le
9	9	290	19.5	244	1	LNRSMC	mannose-binding le
10	10	289.5	19.5	248	1	LNRHMC	mannose-binding le
11	11	287	19.3	248	1	LNRHUP1	mannose-binding le
12	12	285.5	19.2	248	1	LNRHUP1	pulmonary surfacta
13	13	284.5	19.2	248	2	LNRHUP5	pulmonary surfacta
14	14	284	19.1	248	1	LNRHUP6	pulmonary surfacta
15	15	280	18.9	248	1	LNRGPs	pulmonary surfacta
16	16	279	18.8	244	1	LNRMC	mannose-binding le
17	17	277	18.7	248	1	LNRTPS	pulmonary surfacta
18	18	273	18.4	742	2	JC7595	scavenger receptor
19	19	270.5	18.2	239	1	LNRMSA	mannose-binding le
20	20	269.5	18.2	248	2	A48853	pulmonary surfacta
21	21	207	13.9	618	2	S32436	collagen alpha 2(I
22	22	205	13.8	645	2	D90782	probable tail fibe
23	23	205	13.8	645	2	H85642	probable tail fibe
24	24	203	13.7	688	2	A53330	collagen alpha 2(I
25	25	196	13.2	1775	2	A31893	collagen alpha 1(I
26	26	193	13.0	1549	2	I48103	type VII collagen
27	27	192.5	13.0	2944	2	A54849	collagen alpha 1(V
28	28	191.5	12.9	1366	1	CGH251	collagen alpha 2(I
29	29	191	12.9	677	2	S32296	collagen alpha 2(I

A:Title: Primary structure of bovine collectin-43 (CL-43). Comparison with conglutinin a
A:Reference number: A53570; MUID:94216283
A:Accession: A53570
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-301 <LIM>
A:CROSS-references: GB:X75912
R:Holmskov, U.; Teisner, B.; Willis, A.C.; Reid, K.B.; Jensenius, J.C.
J. Biol. Chem. 268, 10120-10125, 1993
A:Title: Purification and characterization of a bovine serum lectin (CL-43) with structure
A:Reference number: A46689; MUID:93252891
A:Accession: A46689
A:Molecule type: protein
A:Residues: 1-27 <HOL>
A:Experimental source: serum
A>Note: sequence extracted from NCBI backbone (NCBIP:131234)
C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
C:Keywords: lectin
F:177-299/Domain: C-type lectin homology <LCH>

Query Match 23.2%; Score 345; DB 2; Length 301;
Best Local Similarity 35.0%; Pred. No. 4e-21;
Matches 86; Conservative 34; Mismatches 98; Indels 28; Gaps 8;

QY 45 SPGPKGDGKGDGEE---GKHGKVGKMGPKGIKGLGMDGDRNTGKTPGKGGK 101
Db 66 SMGPPGTPGKGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 125
QY 102 EKGLLGIPGKGGKAG--TVCDGGRYKVPVGGLDSTARKTSKMKVKNVAGTIRE---- 154
Db 126 LKGDGDPGKGGKAGTSLVVDVTLRQMRNLESEVQRL-----QNVTVQKRAVLFP 178
QY 155 -----TEEKFYIVQENKKNRESITHCIRGGMLAMPKDEAANTLIADYV-AKSGFFRVEI 209
Db 179 DGAAGVEKTEFTAGAVKYSDAEQLEAKGAKGLASPRSAENAEVTLVRAKNK--HAVL 236
QY 210 GVNLEREGQYMETNTPLOYSNNNEGFPD---PYGHEDCVEMLSGGWNTECHLTM 266
Db 237 SMNDISKKGKTYTGTGSL--DYSNAPGEPNNRAKDEGPNCELEIYSDGNWNDIECRE 295
QY 267 YFVCEF 272
Db 296 LVICER 301
:||||
RESULT 3
A45225
pulmonary surfactant protein D precursor - human
N:Alternate names: SP-D
C:Species: Homo sapiens (man)
C:Date: 16-Apr-1999 #sequence_revision 16-Apr-1999 #text_change 22-Jun-1999
C:Accession: A45225; S23434; S24555; S44420; S18382; A56776
R:Crouch, E.; Rust, K.; Velle, R.; Donis-Keller, H.; Grosso, L.
J. Biol. Chem. 268, 2976-2983, 1993
A:Title: Genomic organization of human surfactant protein D (SP-D). SP-D is encoded on
A:Reference number: A45225; MUID:93155122
A:Accession: A45225
A:Molecule type: DNA
A:Residues: 1-375 <CRO>
A:CROSS-references: GB:L05483; GB:L05484; GB:L05485; NID:9292505; PIDN:AAB59450.1; PID:9
A:Experimental source: placenta
A>Note: sequence extracted from NCBI backbone (NCBIP:124316)
R:Lu, J.; Willis, A.C.; Reid, K.B.M.
Biochem. J. 284, 795-802, 1992
A:Title: Purification, characterization and cDNA cloning of human lung surfactant protein
A:Reference number: S23434; MUID:92322003
A:Accession: S23434
A:Molecule type: mRNA
A:Residues: 1-30, T', 32-121, P', 123-179, A', 181-375 <LUI>
A:CROSS-references: EMBL:X65018; NID:934766; PIDN:CAA46152.1; PID:934767
A:Experimental source: lung
A:Accession: S24555

A:Molecule type: protein
A:Residues: 214-234, 'X', 236, 'XX', 239-241 <LUJ2>
R:Hoppe, H.J.; Barlow, P.N.; Reid, K.B.M.
FEBS Lett. 344, 191-195, 1994
A:Title: A parallel three stranded alpha-helical bundle at the nucleation site of col
A:Reference number: S44420; MUID:94244769
A:Accession: S44420
A:Molecule type: mRNA
A:Residues: 202-257 <HOP>
R:Rust, K.; Grosso, L.; Zhang, V.; Chang, D.; Persson, A.; Longmore, W.; Cai, G.Z.; C
Arch. Biochem. Biophys. 290, 116-126, 1991
A:Title: Human surfactant protein D: SP-D contains a C-type lectin carbohydrate recog
A:Reference number: S18382; MUID:91378578
A:Accession: S18382
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 'P', 50-205, 'P', 207-374, 'HF', <RUS>
A:CROSS-references: GB:L05485; NID:9292505
A>Note: corrections to this sequence are reported in reference A56776
R:Crouch, E.; Persson, A.; Chang, D.
Am. J. Pathol. 142, 241-248, 1993
A:Title: Accumulation of surfactant protein D in human pulmonary alveolar proteinosis
A:Reference number: A56776; MUID:93142849
A:Accession: A56776
A>Status: preliminary
A:Molecule type: protein
A:Residues: 46-58, 'P', 60-62, 'E', 64-72, 223-227, 'X', 229-239, 'P', 241-245, 'X', 247-256, 'X'
A:CROSS-references: PIDN:AAB25037.1; PID:q263973; PIDN:AAB25038.1; PID:q263974
A:Experimental source: bronchoalveolar lavage
A>Note: sequence extracted from NCBI backbone (NCBIP:123024, NCBIP:123023); sequence
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower
C:Genetics: This protein is synthesized by alveolar type II cells.
A:Gene: GDB:SFTPD; SFTP4; SP-D
A:CROSS-references: GDB:132674; OMIM:178635
A:Map position: 10q22.2-10q23.1
C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
F:1-20/Domain: blocked amino end; calcium; glycoprotein; hydroxylysine; hydroxyproline;
F:21-375/Product: pulmonary surfactant protein D #status predicted <SIG>
F:21-45/Domain: non-collagenous #status predicted <NCL>
F:46-222/Domain: collagenous #status predicted <COL>
F:223-375/Domain: non-collagenous #status predicted <NC2>
F:254-373/Domain: C-type lectin homology <LCH>
F:90/Binding site: carbohydrate (asn) (covalent) #status predicted
F:281-373, 351-365/disulfide bonds: #status predicted

Query Match 22.7%; Score 337; DB 1; Length 375;
Best Local Similarity 35.2%; Pred. No. 2.4e-20;
Matches 86; Conservative 28; Mismatches 110; Indels 20; Gaps 6;

QY 46 PGPAGDDGKGGPGGEGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGK 96
Db 135 PGPAGGAGKGGGAGPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 194
QY 97 ---KGDKGKGLLIPGKGGKAGT--VDCGRYKRFVGLDSTARKTSKMKVKNY--I 149
Db 195 PGARGPGLKDKGIPGDKGAKGESGLPDVASLRQVQVEALQGVQHLQAASFQYKVELF 254
QY 150 AGIRETEEFYIVQENKKNRESITHCIRGGMLAMPKDEAANTLIAD--YKAGSFRVF 208
Db 255 PNGSGVGEKIFKTAGVFKPTEAQLLCTQAGGLASPRSAENAEALQVLYVAKNE--AAF 312
QY 209 IGAVNLEREGQYMETNTPLOYSNNNEGFPD---PYGHEDCVEMLSGGWNTECHLTM 268
Db 313 LSMDSKTKGKTYTGTGSLV--YSNAPGEPNNRAKDEGPNCELEIYSGNWDIECRE 371
QY 269 VCEF 272
Db 372 VCEF 375
:||||

A>Title: The cDNA cloning of conglutinin and identification of liver as a primary site
A:Reference number: S33235; MUID:93277452
A:Accession: S33235
A:Molecule type: mRNA
A:Residues: 1-172, H¹, I⁷⁴, L¹⁷⁴-217, A¹, V²¹⁹-271, V¹, V²⁷³-371 <LUJ>
A:Cross-references: EMBL:X71774; NID:G395267; PIDN:CAA50665.1; PID:G395268
A:Experimental source: liver
R:Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.
J. Biol. Chem. 266, 2715-2723, 1991
A>Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin family
A:Reference number: A23740; MUID:91131556
A:Accession: A23740
A:Molecule type: protein
A:Residues: 21-209, S¹, S²¹¹-371 <LEE>
R:Kawasaki, N.; Yokota, Y.; Kawasaki, T.
Arch. Biochem. Biophys. 305, 533-540, 1993
A>Title: Differentiation of conglutination activity and sugar-binding activity of conglutinin
A:Reference number: S36879; MUID:93384312
A:Accession: S36879
A:Molecule type: protein
A:Residues: 21-54; 75-86, X¹, H⁸⁸-89, X¹, G⁹¹, X¹, G⁹³-94; 208-209, X¹, I²¹¹-227 <KAW>
A:Experimental source: serum
R:Li, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.
Eur. J. Biochem. 215, 793-799, 1993
A>Title: Structural similarity between lung surfactant protein D and conglutinin. Two different evolutionary relationships
A:Reference number: S35044; MUID:93359905
A:Accession: S35044
A:Molecule type: protein
A:Residues: 75-86, X¹, H⁸⁸-89, X¹, G⁹¹, I¹ <LOJA>
A:Experimental source: lung
R:Young, N.M.; Leon, M.A.
Biochem. Biophys. Res. Commun. 143, 645-651, 1987
A>Title: The carbohydrate specificity of conglutinin and its homology to proteins in the conglutinin family
A:Reference number: A29416; MUID:87184551
A:Contents: annotation
R:Malhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.
Biochem. J. 293, 15-19, 1993
A>Title: Research Communication. Localization of the receptor-binding site in the conglutinin gene exon structure reveals its evolutionary relationship
A:Reference number: S34054; MUID:93319501
A:Contents: annotation
R:Liu, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sasagawa, I.
J. Immunol. 153, 173-180, 1994
A>Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship
A:Reference number: I46010; MUID:94267222
A:Accession: I46010
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-371 <LIO>
A:Cross-references: EMBL:U06860; NID:g507183; PIDN:AAB60624.1; PID:g514256
C:Comment: This protein mediates the agglutination of erythrocytes with antibody and antigen. This protein is a Ca²⁺-dependent serum lectin specific for N-acetylglucosamine 6-phosphate residues.
C:Gene: CGNI
A:Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1
C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
C:Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-371/Product: conglutinin #status predicted <MAT>
F:46-214/Region: collagen-like
F:75-371/Product: conglutinin-N #status predicted <NA>
F:248-369/Domain: C-type lectin homology <CH>
F:63.87.99.135.141.159.162.198.210/Binding site: carbohydrate (Lys) #status expert
F:63.87.99.135.141.159.162.198.210/Modified site: 5-hydroxylysine (Lys) #status expert
F:78.96.108.111.129.132.147.153.171.195/Modified site: 4-hydroxyproline (Pro) #status expert
F:337/Binding site: carbohydrate (Asn) #status experimental

Query Match 22.1%; Score 328.5; DB 1; Length 371;
Best Local Similarity 29.5%; Pred. No. 1.2e-19;
Matches 85; Conservative 35; Mismatches 95; Indels 73; Gaps 10;

QY 46 PGPKGDDGEKDP-----GEERKHGVGRMGPGIKGLGDMDRGNTGK-----TGP 93
||||| : | : ||||| : | : ||||| : | : ||||| : | : ||||| : | :

[illegible]

A:Cross-references: EMBL:X15954; NID:g34480; PIDN:CAA34079.1; PID:g1212951
A:Accession: A34978
A:Molecule type: protein
A:Residues: 'X',22-24,'X',26,'X',28-31,'X',33-34,'X',36,'XXXX',41-50 <TAYZ>
R:Rzekowits, R.A.B.; Day, L.E.; Herman, G.A.
J. Exp. Med. 167, 1034-1046, 1988
A:Title: A human mannose-binding protein is an acute-phase reactant that shares sequence
A:Reference number: J10027; MUID:88171281
A:Accession: J10027
A:Molecule type: mRNA
A:Residues: 1-2,'C',4,'T',8,'S',10-57,'R',59-60,'G',63-106,'PGCLRK',113,'SSANNGTYQ',1
R:Kurata, H.; Sannoh, T.; Kozutsumi, Y.; Yokota, Y.; Kawasaki, T.
J. Biochem. 115, 1148-1154, 1994
A:Title: Structure and function of mannan-binding proteins isolated from human liver and
A:Reference number: JX0319; MUID:95073978
A:Accession: JX0319
A:Molecule type: protein
A:Residues: 1-248 <KUR1>
A:Accession: PC2188
A:Molecule type: mRNA
A:Residues: 1-20 <KUR2>
A:Experimental source: liver and serum
C:Comment: Mannose-binding lectins are opsonins that are important in host defense against
C:Comment: This protein is a Ca2+-requiring animal lectin specific for mannose and N-acetyl
C:Genetics:
A:Gene: GDB:MBL
A:Cross-references: GDB:120167; OMIM:154545
A:Map position: 10q11.2-10q11.2
A:Introns: 63/1; 102/1; 125/1
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: acute phase; calcium binding; endoplasmic reticulum; Golgi apparatus; hydrox
F:1-20/Domain: signal sequence status predicted <SG>
F:21-248/Product: mannose-binding lectin #status experimental <MAT>
F:42-99/Region: collagen-like
F:128-244/Domain: C-type lectin homology <LCH>
F:47,73,79,82,88/Modified site: 4-hydroxyproline (Pro) (partial) #status experimental

[illegible]

RESULT 11
LNHUPI
pulmonary surfactant protein A precursor (clone 1A) - human
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: B25720
R:Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman, L.;
J. Biol. Chem. 261, 9029-9033, 1986
A>Title: Isolation and characterization of cDNA clones for the 95-kDa pulmonary surfactant

A:Reference number: A25720; MUID:86250832
A:Accession: B25720
A:Molecule type: mRNA
A:Residues: 1-248 <FLO>
A:Cross-references: GB:K03475
A:Note: part of the sequence was confirmed by protein sequencing
A:Note: the amino end of the mature protein, which was not identified, is partially a
A:Note: clones corresponding to two different proteins were sequenced. Cotranslational
C:Genetics:
A:Gene: GDB:SFTPAL; SFTPL; SP-A; SP-A1
A:Cross-references: GDB:119593; OMIM:178630
A:Map position: 10q22-10q23
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; g1
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F:127-245/Domain: C-type lectin homology <LGH>
F:21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted
F:307,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status p
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.3%; Score 287; DB 1; Length 248;
Best Local Similarity 28.2%; Pred. No. 2e-16;
Matches 75; Conservative 44; Mismatches 117; Indels 30; Gaps 9;

QY 14 ILLVFLLIQISLGLDIDSRPTAEVCATHTISPGKGDGKDPGEEGKHGKVGRMGPK 73

db 6 1A1T11MAASGACEV----KNVVC---SDCIPTCTCSUCI DCPNCPNATKCCNCCD 57

[illegible]

QY 74 GIKGELGDM---GDRGNIGKTGPIGKKGDKGEKGLLGIPGEKGAGTVDCCGRYKFFVG 129

Db 58 GPMGPPGETPCPPGNNGLPAGVGPGEKGE-----PGERGPPGLPAHLDE----- 105

QY 130 QLDISIARLKTSMKFVKNVIA---GIRETEEFYIVQEEKNYRESLTHCRGGMLAMP 186

db 106 ELQATIHDEPHOITRECAI SIACSTIMVCEKVESNCSSTEDATOFKACBACIATINS 168

[illegible]

QY 187 KDEAANTLIADYVAKSGFFRVFIGVNDLEREGQYMFDTNTPLQYNSNNWEEGEPSPDYGHE 246

Db 166 RNPEENEAIASFVKKNTY -AYVGLTEGSPGDFRYS DGTVP -NYTNWYRGEPAG -RGKE 222

QY 247 DCVEMLSSGRNDTECHLTMVFVCEF 272

: |||| : |:||| | : :|||
223 QGVEMVWFCQENIDNCIYCDITMOTN 240

223 Q06TII DGQWNRNCLISREICEF 248

RESULT 12

LNHUPS

N:Alternate names: alveolar proteinosis protein; pulmonary surfactant precursor (genomic clone) - human

C;Species: Homo sapiens (man)

C;Accession: A24622; A43628
C;date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

R;White, R.T.; Damm, D.; Miller, J.; Spratt, K.; Schilling, J.; Hawgood, S
Nature 317 361-363 1995

A>Title: Isolation and characterization of the human pulmonary surfactant a

A;Reference number: A24622; MUID:86014366
A;Accession: A24622

A: Molecule type: DNA

A;Residues: I-248 <WHI>
A;Cross-references: GB:M30838: NID:σ190564: PIDN:AAA36510.1: PTD:σ190565

A>Note: the sequence in GenBank entry HUMPSAP, release 109.0, (PID:g190565) differs from the sequence in GenBank entry HUMPSAP, release 109.0, (PID:g190565) by 4 nucleotides.

R. Haagsman, H.P.; White, R.T.; Schilling, J.; Lau, K.; Benson, B.J.; Golden, A. Note: Four nucleotide differences, producing amino acid differences at positions 10, 11, 12, and 13, were found in the coding region of the *hprt* gene in the R. Haagsman, H.P.; White, R.T.; Schilling, J.; Lau, K.; Benson, B.J.; Golden, A. Note: Four nucleotide differences, producing amino acid differences at positions 10, 11, 12, and 13, were found in the coding region of the *hprt* gene in the

Am. J. Physiol. 257, L421-L429, 1989

A, Reference number: A43628; MWID: 90119861

A:Accession: A43628
A:Molecule type: protein

A:Residues: 143-150;220-240;243-248 <HAA>

C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins

[illegible]

[illegible]

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	1484	100.0	277	20	AAV25518	Human collectin pr
2	1472	99.2	277	20	AAV41958	Human PRO702 prote
3	1472	99.2	277	21	AAB44254	Human PRO702 (UNQ3
4	1472	99.2	277	22	AAU29073	Human PRO polypept
5	716.5	48.3	271	21	AAB24075	Human PRO1182 prot
6	716.5	48.3	271	21	AAV79510	Human carboxylate ✓
7	716.5	48.3	271	21	AAV66738	Membrane-bound pro
8	716.5	48.3	271	22	AAV65261	Human PRO1182 (UNQ
9	696.5	46.9	292	22	AAW78426	Human protein SEQ
10	693.5	46.7	263	22	AAW23731	Human EST encoded
11	623.5	42.0	268	22	AAW78427	Human protein seq

RESULT

Claim 1; Page

This invention describes the isolation and characterisation of a novel

CC human collectin protein and its encoding polynucleotide. The human
 CC collectin exhibits antibacterial and antiviral activity and can be used
 CC as an agent for the treatment of human bacterial and viral infections.
 CC This sequence represents the novel human collectin.

XX Sequence 277 AA;

Query Match 100.0%; Score 1484; DB 20; Length 277;
 Best/Local Similarity 100.0%; Pred. No. 4.7e-142;
 Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNGFASLRNQIFILLVLEFLOIQLDIDSRPTAEVCAHTTISPGKGDGKGDGPE 60
 Db 1 mngfasllrrngfillvllfllqslgldidsrptaevcahtisppgkddgkdpge 60
 QY 61 EKHGKVGMPGPKIGELGDMGRNGTGTGPIGKKGDKGKGLLIPGKKGKAGTVCD 120
 Db 61 egkhgkvgrmpgpkigeldgmdgrngtgtgpi gkkgdkgkglilgipgkkgagtvc 120
 QY 121 CGYRKFGQLDISARLTKSMKFKVKNVIAGIRETEKYYIVQEKKNYRESLTHCRIRG 180
 Db 121 cgyrkfvgqldisarltskmskfkvnviagireteekfyivqekknysrslthcrirg 180
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRVFIGNDLEREGQYMTDNTPLQYSNWNEGEPS 240
 Db 181 gmlampkdeaanliadyvaksgffrvfignvdleregymftdntplqysnwnegeps 240
 QY 241 DPYGHEDCVEMLSGRWNTECHLMFVCEFTKKKK 277
 Db 241 dpyghedcvemlssgrwndtechltmyfvcfeikkk 277

RESULT 2

AAAY1698
 ID AAY41698 standard; Protein; 277 AA.

AC AAY41698;

XX 07-DEC-1999 (first entry)

XX Human PRO702 protein sequence.

XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 KW secreted protein; transmembrane protein.

XX Homo sapiens.

XX WO9946281-A2.

XX 16-SEP-1999.

XX 08-MAR-1999; 99WO-US05028.

XX 10-MAR-1998; 98US-0077450.

XX 11-MAR-1998; 98US-0077632.

XX 11-MAR-1998; 98US-0077641.

XX 11-MAR-1998; 98US-0077649.

XX 12-MAR-1998; 98US-0077731.

XX 13-MAR-1998; 98US-0078004.

XX 17-MAR-1998; 98US-0040220.

XX 20-MAR-1998; 98US-0078910.

XX 20-MAR-1998; 98US-0078936.

XX 20-MAR-1998; 98US-0078939.

XX 25-MAR-1998; 98US-0079294.

XX 26-MAR-1998; 98US-0079656.

XX 27-MAR-1998; 98US-0079663.

XX 27-MAR-1998; 98US-0079664.

XX 27-MAR-1998; 98US-0079689.

XX 27-MAR-1998; 98US-0079728.

XX 27-MAR-1998; 98US-0079786.

PR 30-MAR-1998; 98US-0079920.
 PR 30-MAR-1998; 98US-0079923.
 PR 31-MAR-1998; 98US-0080105.
 PR 31-MAR-1998; 98US-0080107.
 PR 31-MAR-1998; 98US-0080165.
 PR 31-MAR-1998; 98US-0080194.
 PR 01-APR-1998; 98US-0080327.
 PR 01-APR-1998; 98US-0080328.
 PR 01-APR-1998; 98US-0080333.
 PR 01-APR-1998; 98US-0080334.
 PR 08-APR-1998; 98US-0081049.
 PR 08-APR-1998; 98US-0081070.
 PR 08-APR-1998; 98US-0081071.
 PR 09-APR-1998; 98US-0081195.
 PR 09-APR-1998; 98US-0081203.
 PR 09-APR-1998; 98US-0081229.
 PR 15-APR-1998; 98US-0081817.
 PR 15-APR-1998; 98US-0081838.
 PR 15-APR-1998; 98US-0081952.
 PR 15-APR-1998; 98US-0081955.
 PR 21-APR-1998; 98US-0082568.
 PR 21-APR-1998; 98US-0082569.
 PR 22-APR-1998; 98US-0082700.
 PR 22-APR-1998; 98US-0082704.
 PR 22-APR-1998; 98US-0082804.
 PR 23-APR-1998; 98US-0082767.
 PR 23-APR-1998; 98US-0082796.
 PR 27-APR-1998; 98US-0083336.
 PR 28-APR-1998; 98US-0083322.
 PR 29-APR-1998; 98US-0083392.
 PR 29-APR-1998; 98US-0083495.
 PR 29-APR-1998; 98US-0083496.
 PR 29-APR-1998; 98US-0083499.
 PR 29-APR-1998; 98US-0083500.
 PR 29-APR-1998; 98US-0083545.
 PR 29-APR-1998; 98US-0083554.
 PR 29-APR-1998; 98US-0083558.
 PR 29-APR-1998; 98US-0083559.
 PR 30-APR-1998; 98US-0083742.
 PR 05-MAY-1998; 98US-0084366.
 PR 06-MAY-1998; 98US-0084414.
 PR 06-MAY-1998; 98US-0084441.
 PR 07-MAY-1998; 98US-0084598.
 PR 07-MAY-1998; 98US-0084600.
 PR 07-MAY-1998; 98US-0084627.
 PR 07-MAY-1998; 98US-0084637.
 PR 07-MAY-1998; 98US-0084639.
 PR 07-MAY-1998; 98US-0084640.
 PR 07-MAY-1998; 98US-0084643.
 PR 13-MAY-1998; 98US-0085323.
 PR 13-MAY-1998; 98US-0085338.
 PR 13-MAY-1998; 98US-0085339.
 PR 15-MAY-1998; 98US-0085573.
 PR 15-MAY-1998; 98US-0085579.
 PR 15-MAY-1998; 98US-0085580.
 PR 15-MAY-1998; 98US-0085582.
 PR 15-MAY-1998; 98US-0085689.
 PR 15-MAY-1998; 98US-0085697.
 PR 15-MAY-1998; 98US-0085700.
 PR 15-MAY-1998; 98US-0085704.
 PR 18-MAY-1998; 98US-0086023.
 PR 22-MAY-1998; 98US-0086392.
 PR 22-MAY-1998; 98US-0086414.
 PR 22-MAY-1998; 98US-0086430.
 PR 22-MAY-1998; 98US-0086486.
 PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.

(GETH) GENENTECH INC.

XX

PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 XX WPI; 1999-551358/46.
 DR N-PSDB; AAZ33973.
 XX
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 PS Claim 12; Fig 37; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AAZ33891 to
 CC AAZ34338, and AAZ41685 to AAZ41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 20; Length 277;
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNGFASLLRNQFILLVLLQIQTSLGLDIDSRPTAEVCATHTISPGKGDGEGKDPGE 60
 Db 1 mngfasllrnqfllvlllqislgldidspptaevcathtispgkdgdekdpge 60
 QY 61 EGKHGKVGMPGKIGKELGMDGRNIGKTGPIGKKGDKGKGLGIPGKAGTVC 120
 Db 61 egkhgkvgmpgkigkeldgmdgrnigktgpihgkkgdkgkglgipgekagtvc 120
 QY 121 CGRYKRVGQDLSIARLKTSMKFKVKNVIAGIRETEKFFYIVQEKNYRESLTHCRIRG 180
 Db 121 cgryrkfvqgldislarlktsmkfkvnviagireteekffyyivqeknyreslthcrirg 180
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFRVIGVNDLEREGQYMETDNTPLQYSNWNEGEP 240
 Db 181 gmlampkdeaanliadyvaksfffrfrvignvndlereregymstdntplqysnwnegeps 240
 QY 241 DPYGHEDCVEMLSGRWNTDTECHLTMYFVCFEIKKK 277
 Db 241 dpyghedcvemlsgrwndtchltmyfvcefikkk 277

RESULT 3
 AAB44254
 ID AAB44254 standard; Protein; 277 AA.

AC AAB44254;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human PR0702 (UNQ366) protein sequence SEQ ID NO:97.
 XX
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
 KW expressed sequence tag; detection; cancer.
 XX
 OS Homo sapiens.
 XX
 FN W0200053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.

PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 XX WPI; 2000-611443/58.
 DR N-PSDB; AAZ78480.

Novel PRO polypeptides and polynucleotides used in detection methods,
 to target bioactive molecules to specific cells, and to modulate
 cellular activities -

Claim 12; Fig 37; 636pp; English.

CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 sequence tag) sequences which encode secreted or transmembrane PRO
 polypeptides. The PRO polynucleotides and polypeptides have cytostatic
 activity. The polynucleotides and polypeptides can be used for detecting
 the presence of PRO polypeptides in samples, for linking bioactive
 molecules to cells and for modulating biological activities of cells,
 using the polypeptides for specific targeting. The polypeptide targeting
 can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 the isolation of the PRO polynucleotide sequences.

Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 21; Length 277;
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNGFASLLRNQFILLVLLQIQTSLGLDIDSRPTAEVCATHTISPGKGDGEGKDPGE 60
 Db 1 mngfasllrnqfllvlllqislgldidspptaevcathtispgkdgdekdpge 60
 QY 61 EGKHGKVGMPGKIGKELGMDGRNIGKTGPIGKKGDKGKGLGIPGKAGTVC 120
 Db 61 egkhgkvgmpgkigkeldgmdgrnigktgpihgkkgdkgkglgipgekagtvc 120
 QY 121 CGRYKRVGQDLSIARLKTSMKFKVKNVIAGIRETEKFFYIVQEKNYRESLTHCRIRG 180
 Db 121 cgryrkfvqgldislarlktsmkfkvnviagireteekffyyivqeknyreslthcrirg 180
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFRVIGVNDLEREGQYMETDNTPLQYSNWNEGEP 240
 Db 181 gmlampkdeaanliadyvaksfffrfrvignvndlereregymstdntplqysnwnegeps 240
 QY 241 DPYGHEDCVEMLSGRWNTDTECHLTMYFVCFEIKKK 277
 Db 241 dpyghedcvemlsgrwndtchltmyfvcefikkk 277

RESULT 4
 AAU29073
 ID AAU29073 standard; Protein; 277 AA.
 XX
 AC AAU29073;
 DT 18-DEC-2001 (first entry)
 DE Human PRO polypeptide sequence #50.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 PN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX
 PE 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194677P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000US-0644848.
 PR 08-NOV-2000; 2000WO-US33328.
 PR 01-DEC-2000; 2000WO-US30952.
 PR 20-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PU, Gurney AL;
 PI Fan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-602746/68.
 DR N-PSDB; AAS45974.
 XX

PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 XX to screen for modulators of the compounds -
 PS Claim 11; Fig 100; 774pp; English.
 XX
 CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX Sequence 277 AA;
 SQ

Query Match 99.2%; Score 1472; DB 22; Length 277;
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNGFASLLRRNQFILLVFLQLQSLGLDIDSRPTAEVCATHTISPGKGGDGEKGPGE 60
 Db 1 mngfasllrrnqfllvflqlqslgldidsrptaevcathitspgkggdgekdpge 60
 QY 61 B3K5KVGWGPCKIGKELGDMGDRNICKTGPCKYKGDGKGLLGPBKGKAGTVCD 120
 Db 61 egkhgkvgrmgpckigkeldmgdrgncktpckykdgdkglgllgpbgkagtvcd 120
 QY 121 CGRYKRVGQLDISIARLKTSMFVKVNIAGIRTEKEFYIVOREKNYRESLTHCRIRG 180
 Db 121 cgryrkfvvgldisarlktsmfknvniagireteekfyivoreknyreslthcrrirg 180
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRVFIGVNDLEREGQYMETDNTPLQNSWNNEGEP 240
 Db 181 gmlampkdeaanltiadyvaksgffrvfigvndlerregqymstdntplqnswnnegeps 240
 QY 241 DPYGHEDCVEMLSGRWNTDTECHLTMYFVCFEIRKKK 277
 Db 241 dpyghedcvemlssgrwntdtechlmyfvcfeikkk 277

RESULT 5
 AAB24075
 ID AAB24075 standard; Protein; 271 AA.
 XX
 AC AAB24075;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human PRCL182 protein sequence SEQ ID NO:51.
 KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
 KW proliferation; tumorigenesis; identification; cancer; cytostatic;
 KW nontropic; neuroprotective; antiinflammatory; immunosuppressive;
 KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
 KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
 KW hypothalamic disorder; glandular disorder; macrophagal disorder;
 KW epithelial disorder; stromal disorder; blastocoelec disorder;
 XX inflammatory disorder; immunologic disorder.
 XX Homo sapiens.
 OS
 XX
 PN WO2000053755-A2.

PR	09-JUN-1998	9805-00865508
PR	10-JUN-1998	9805-00867222
PR	10-JUN-1998	9805-00887334
PR	10-JUN-1998	9805-00887334
PR	10-JUN-1998	9805-00887338
PR	10-JUN-1998	9805-00887401
PR	10-JUN-1998	9805-00887410
PR	10-JUN-1998	9805-00887421
PR	10-JUN-1998	9805-00888110
PR	10-JUN-1998	9805-00888610
PR	10-JUN-1998	9805-00888624
PR	10-JUN-1998	9805-00888625
PR	10-JUN-1998	9805-00888626
PR	11-JUN-1998	9805-00888658
PR	11-JUN-1998	9805-00888661
PR	11-JUN-1998	9805-00888663
PR	11-JUN-1998	9805-00888676
PR	12-JUN-1998	9805-00888762
PR	12-JUN-1998	9805-00890390
PR	16-JUN-1998	9805-00894103
PR	16-JUN-1998	9805-00894444
PR	16-JUN-1998	9805-00895112
PR	16-JUN-1998	9805-00895114
PR	17-JUN-1998	9805-00895332
PR	17-JUN-1998	9805-00895338
PR	17-JUN-1998	9805-00895358
PR	17-JUN-1998	9805-00895599
PR	17-JUN-1998	9805-00896600
PR	17-JUN-1998	9805-00896533
PR	18-JUN-1998	9805-00898031
PR	18-JUN-1998	9805-00899007
PR	19-JUN-1998	9805-00899008
PR	19-JUN-1998	9805-00899477
PR	19-JUN-1998	9805-00899488
PR	22-JUN-1998	9805-00899552
PR	22-JUN-1998	9805-00902446
PR	22-JUN-1998	9805-00902552
PR	23-JUN-1998	9805-00902554
PR	23-JUN-1998	9805-00903449
PR	23-JUN-1998	9805-00903555
PR	24-JUN-1998	9805-00904272
PR	24-JUN-1998	9805-00904293
PR	24-JUN-1998	9805-00904311
PR	24-JUN-1998	9805-00904335
PR	24-JUN-1998	9805-00904444
PR	24-JUN-1998	9805-00904455
PR	24-JUN-1998	9805-00904611
PR	24-JUN-1998	9805-00904772
PR	24-JUN-1998	9805-00905335
PR	24-JUN-1998	9805-00905338
PR	24-JUN-1998	9805-00905440
PR	24-JUN-1998	9805-00905557
PR	25-JUN-1998	9805-00906776
PR	25-JUN-1998	9805-00906778
PR	25-JUN-1998	9805-00906862
PR	25-JUN-1998	9805-00906868
PR	26-JUN-1998	9805-00906930
PR	01-JUL-1998	9805-00913358
PR	01-JUL-1998	9805-00913500
PR	01-JUL-1998	9805-00913644
PR	02-JUL-1998	9805-00914778
PR	02-JUL-1998	9805-00914866
PR	02-JUL-1998	9805-00915119
PR	02-JUL-1998	9805-00916236
PR	02-JUL-1998	9805-00916238
PR	02-JUL-1998	9805-00916333
PR	02-JUL-1998	9805-00916446
PR	02-JUL-1998	9805-00916773
PR	07-JUL-1998	9805-00919782
PR	07-JUL-1998	9805-00919782

CC will also be useful for the preparation of PRO polypeptides, especially by recombinant techniques.

XX sequence 271 AA;

Query Match 48.3%; Score 716.5; DB 21; Length 271;
Best Local Similarity 47.4%; Pred. No. 3e-64;
Matches 129; Conservative 60; Mismatches 76; Indels 7; Gaps

QY 8 LRRNQFILLVLLIQISGLGIDSRPTA-EVCATHITSPGKSGDDGKGPFGEGKHGK 66
DB 1 mrglaalvgvlslafslslpshpdpagddacsqvlvpqllkgdagekgd---kgaqgr 57
QY 67 VGRMGPKGIGELGDMGDRGNI---GKTGPTGKKGKGLGILGIPGKKGAGTVCDCGR 123
DB 58 pgrvgtgkqmgdkgkqsvgrhkiplgskgkqsdgipgpgnggeglpccsq 117
QY 124 YRFVQGOLDISARLKTSMFKVKNVIAGIRETEKFYYIYQEEKNYRESLTHCRIGGML 183
DB 118 lrkaigemdngvqsltselkfklnavavreteskiyllvkeekryadaqlscqgrggtl 177
QY 184 AMPKDAEANTLLADYAKSGEFFRVFVGVNDLERGQVMTDNTPLQNTSYNNWGEFSDPY 243
DB 178 smpkdeanglnaaylaqaglarvfigindlekegafvysdhspmrtfkvrsgpepnay 237
QY 244 GHEDCVEMLSGGRWNDTECHLTMYFVCEFIKK 275
DB 238 deedcvemvasggnwadvachttmymfmcfdke 269

RESULT 8
ID AAB65261
XX ID AAB65261 standard; Protein; 271 AA.
XX AAB65261;
DT
DT (first entry)
DE Human PRO1182 (UNQ596) protein sequence SEQ ID NO:357.
DE
KW Human; secreted and transmembrane protein; PRO; cytosolic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW diagnostic assay.
XX
XX Homo sapiens.
OS
PN WO2000073454-A1.
PD
PD 07-DEC-2000.
PF 30-MAR-2000; 2000WO-US08439.
PX 02-JUN-1999; 99WO-US12252.
PX 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.

PR 24-FEB-2000; 2000WO-US05004.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 20-MAR-2000; 2000WO-US07377.
 XX (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DJ;
 PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
 PI Grimaldi CJ, Gurney AL, Kljavin LJ, Napier MA, Pan J, Paoni NF;
 PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
 PI Zhang Z;
 XX
 XX WPI: 2001-032160/04.
 DR N-PSDB; AAF44230.
 XX
 XX PRO polynucleotides used to produce polypeptides used to target
 PT bioactive molecules such as toxins, radiolabels or antibodies, to
 PT specific cells, to cause targeted cell death -
 XX
 PS Claim 12; Fig 252; 935pp: English.
 XX
 XX The present invention describes human secreted and transmembrane PRO
 CC proteins. The PRO proteins have cytosolic activity. The PRO proteins
 CC can be used for targeted delivery of bioactive molecules, such as
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
 CC sequences and their fragments, can be used as hybridisation probes, in
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA
 CC and DNA. They may also be used to produce transgenic animals which are
 CC used to develop and screen therapeutically useful reagents. The PRO
 CC nucleotide and protein sequence can be used for tissue typing and in
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
 CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
 CC AAF5154 to AAF5300 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.
 XX
 XX Sequence 271 AA;
 SQ
 Query Watch 48.3%; Score 716.5; DB 22; Length 271;
 Best Local Similarity 47.4%; Pred. No. 3e-64;
 Matches 129; Conservative 60; Mismatches 76; Indels 7; Gaps 3;
 QY 8 LERNQFLLVLLQLQSLGLDIDSRPTA-EVCATHITSPGKDGEGKDPGEGKHGK 66
 Db 1 mrgnlavglvllslafslpsghpdpagddacsvqlvpqlkgdagekgd---kgapgr 57
 QY 67 VGRMGPKGKIGELGMDGRGNI---GKTGPICKGDKGKGLLGIPEGKRGAGTVCDCGR 123
 Db 58 pgrvgtgkdgmgkgkgvgrhkgipgskgkgsdgidgppgngpgeplpcecsq 117
 QY 124 YKXFGOLDISTARLKTSMKFKVKNVIAGIRETEEFYIVQEEKNYRSLTHCRIRGGM 183
 Db 118 lrlkaigmdqvsqsltselkfnknagvreteskiyllylVkeekyadaqlscgrggtl 177
 QY 184 AMPKDEAANTLIADYVAKSGFFRVFGVNDLERSGQYMFDTNPLQNSNNWEGPSDPY 243
 Db 178 smpkdeaanlmaaylaqaglarvfigindlekafvysnshpmrtfnkwrsgpuna 237
 QY 244 GHEDCVMLSSGRWNNDTECHTMYVCFEIKK 275
 Db 238 deedeemvasgfgndvachttmyfmcfdke 269
 RESULT 9
 AAM78426
 ID AAM78426 standard; Protein; 292 AA.
 XX
 XX AAM78426;
 XX
 DT 06-NOV-2001 (first entry)
 XX

DE Human protein SEQ ID NO 1088.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 XX WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue A, Yang Y, Wejhrman T, Goodrich R;
 XX
 XX WPI: 2001-476283/51.
 DR N-PSDB; AAK51559.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 3318-3319; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 292 AA;
 Query Match 46.9%; Score 696.5; DB 22; Length 292;
 Best Local Similarity 46.2%; Pred. No. 3.6e-62;
 Matches 126; Conservative 58; Mismatches 80; Indels 9; Gaps 3;
 QY 3 GFASLLRRNQFILLVLLQLQSLGLDIDSRPTA---EVCATHITSPGKDGEGKDPG 59
 Db 23 gvlpalrmrgnlavglvllslafslpsghpdpagddacsvqlvpqlkgdagekgd-- 80
 QY 60 EECKHGKVGMRGMPKIGELGMDGRGNI---GKTGPICKGDKGKGLLGIPEGKRGAG 116
 Db 81 -kgapgrprvgtgkdgmgkgkgvgrhkgipgskgkgsdgidgppgngpgepg 139
 QY 117 TVDCGGRYKFEVQGLDISTARLKTSMKFKVKNVIAGIRETEEFYIVQEEKNYRSLTHC 176
 Db 140 lpcecsqirkaigmdqvsqsltselkfnknagvreteskiyllylVkeekyadaqlsc 199
 QY 177 RINGMLAMPKDEAANTLIADYVAKSGFFRVFGVNDLERSGQYMFDTNPLQNSNNWNE 236


```
XX SQ Sequence 268 AA;
Query Match 42.0%; Score 623.5; DB 22; Length 268;
Best Local Similarity 42.2%; Pred. No. 8e-55;
Matches 114; Conservative 53; Mismatches 76; Indels 27; Gaps 2;
QY 3 GFASLLRNQILVFLVLIQISGLDIDSPTA---EVCATHHTISPGKDDGKGDPG 59
DQ 23 gvlpalmrgnlaivglislaflslpsghpapgddacsqilvpglkdgadegkdkg 82
QY 60 EEGKGGKVMGPKGKINGELGMDGRGNCKTGPICKKDGKGEKGLLIGPCGKAGIVC 119
DQ 83 aprrgrv-----gptgkgekgdsgdipppngpgepglpc 118
QY 120 DGRYKRVGQDLSIARLKTSMKFNKVIAGIRETEEFYIVQEEKNYRESLTHCRIR 179
DQ 119 ecsqirkaigemdngvqsltselkfkknagvreteskiyllvkeekryadaqlscgr 178
QY 180 GGLAMPKDEAANTLIADYVAKSGFRFVIGVNDLEREGQYMTDNTPLQNSNNNEGEP 239
DQ 179 ggtlmpkdeanglnaaylaqagiaryfigindlekeafydsghpsmtfnkvrsgpe 238
QY 240 SDPYGHEDCVEMLSGRWMDTECHLTMYEV 269
DQ 239 maydeedcvenvasggnwadvactctmyfm 268

RESULT 12
ABG15021
ID ABG15021 standard; Protein; 222 AA.
XX
AC ABG15021;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15012.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR N-PSDB; AAS79208.
XX
WPI: 2001-639362/73.
XX
N-PSDB; AAS79208.
XX
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.
XX
Claim 20; SEQ ID No 45380; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
```

```
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 222 AA;
```

```
Query Match 25.3%; Score 376; DB 22; Length 222;
Best Local Similarity 36.3%; Pred. No. 7.7e-30;
Matches 77; Conservative 41; Mismatches 68; Indels 26; Gaps 2;
QY 28 LDIDSRPTAEVCAHTISPGKDGKGEKGPGEKKGKGVGRMGPKGIKGBELGMDGDRGN 87
DQ 11 ldilsrpgddacsqilvpglkdgadegkdkgaprgprv----- 50
QY 88 IGKTGPIGKKDGKGEKGLLIGPCGKAGTVCDCGRYKRVGQDLSIARLKTSMKFNK 147
DQ 51 ----gptgkgekgdsgdipppngpgepglpoecsqrkaigemdngvqsltselkfk 106
QY 148 VIAGIRETEEFYIVQEEKNYRESLTHCRIRGGGLAMPKDEAANTLIADYVAKSGGFFRV 207
DQ 107 avagvreteskiyllvkeekypadpplswqgrggtlsmadqavhglkgylgqarlgrg 166
QY 208 FIGVND--LEREGQYMTDNTPLQNSNNNEG 237
DQ 167 figihprwrraplllftlsplrsfxkdg 198
```

```
RESULT 13
AA777990
ID AA777990 standard; Protein; 375 AA.
XX
AC AA777990;
XX
DT 20-JUN-2000 (first entry)
XX
DE Human SP-D amino acid sequence.
XX
KW Collectin; human; antibacterial; antiviral; SP-D.
XX
OS Homo sapiens.
XX
PN WO200011161-A1.
XX
PD 02-MAR-2000.
XX
PF 24-AUG-1999; 99WO-JP04552.
XX
PR 24-AUG-1998; 98JP-0237611.
XX
PA (FUSO ) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
WPI: 2000-224696/19.
XX
New collectin of human origin having antibacterial and antiviral
activity, and gene encoding it useful for production of transgenic
PT animals and of antibodies for screening potential drug molecules -
XX
Disclosure; Fig 2-3; 106pp; Japanese.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:52 ; Search time 15.79 Seconds
(without alignments)
679.247 Million cell updates/sec

Title: US-09-600-932-2

Perfect score: 1484

Sequence: 1 MNGFASLLRRNQFLLEFL.....NDTECHLWYFVCFIKKKK 277

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349.5	23.6	369	1	PSPD_BOVIN
2	344	23.2	301	1	CA13_CHICK
3	340	22.9	374	1	PSPD_MOUSE
4	337	22.7	375	1	PSPD_HUMAN
5	333	22.4	374	1	PSPD_RAT
6	328.5	22.1	371	1	CONG_BOVIN
7	305.5	20.6	247	1	PSPA_RABIT
8	290	19.5	238	1	MABA_RAT
9	290	19.5	244	1	MABC_MOUSE
10	289.5	19.5	248	1	MABC_HUMAN
11	285.5	19.2	249	1	PSPA_PIG
12	284.5	19.2	249	1	MABC_BOVIN
13	284	19.1	248	1	PSPA_HUMAN
14	280	18.9	248	1	PSPA_CANEA
15	279	18.8	244	1	MABC_RAT
16	277	18.7	248	1	PSPA_RAT
17	270.5	18.2	239	1	MABA_MOUSE
18	270	18.2	247	1	PSPA_CAVPO
19	269.5	18.2	248	1	PSPA_MOUSE
20	207	13.9	689	1	CA29_HUMAN
21	196	13.2	1775	1	CA14_DROME
22	192.5	13.0	2944	1	CA17_HUMAN
23	191.5	12.9	1366	1	CA21_HUMAN
24	191	12.9	1022	1	CA26_CHICK
25	189.5	12.8	1018	1	CA26_HUMAN
26	187.5	12.6	518	1	MTCO_MOUSE
27	187	12.6	1029	1	CA26_MOUSE
28	186	12.5	1025	1	CA26_MOUSE
29	186	12.5	1650	1	CA2B_MOUSE
30	185.5	12.5	674	1	CA1A_CHICK
31	184.5	12.4	1362	1	CA21_CHICK
32	183.5	12.4	1364	1	CA21_BOVIN
33	183	12.3	1838	1	CA15_HUMAN

34 182 12.3 1262 1 CA13_CHICK
35 181.5 12.2 1019 1 CA16_CHICK
36 181.5 12.2 1603 1 CA1F_HUMAN
37 180.5 12.2 1366 1 CA21_CANEA
38 180.5 12.2 1496 1 CA25_HUMAN
39 179.5 12.1 1028 1 CA16_HUMAN
40 179 12.1 520 1 MRCO_HUMAN
41 178.5 12.0 1372 1 CA21_MOUSE
42 178.5 12.0 1669 1 CA14_HUMAN
43 177.5 12.0 483 1 MRCO_MESAU
44 177.5 12.0 1418 1 CA12_HUMAN
45 177 11.9 547 1 CAFL_EPHMU

P12105 gallus gall
P20785 gallus gall
Q07092 homo sapien
O46392 canis famil
P05997 homo sapien
P12109 homo sapien
Q9uew3 homo sapien
Q01149 mus musculu
P02462 homo sapien
Q9wub9 mesocricetu
P02458 homo sapien
P18856 ephydatia m

ALIGNMENTS

RESULT 1
PSPD_BOVIN
ID PSPD_BOVIN STANDARD; PRT: 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
RC TISSUE=Lung;
RX MEDLINE=93170856; PubMed=8436402;
RA Lim B.L., Lu J., Reid K.B.M.;
RT "Structural similarity between bovine conglutinin and bovine lung
surfactant protein D and demonstration of liver as a site of
RT synthesis of conglutinin.";
RL Immunology 78:153-165(1993).
CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE
EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X75911; CAA53510.1; -
DR PIR: S33603; S33603.
DR HSP: P35247; 1808.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; Lectin_c.
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF00059; Lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LLECTIN_2; 1.
DR Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
Signal; Lactin; Collagen; Repeat; Coiled coil.
KW

CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
 CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
 CC -|- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -|- SUBCELLULAR LOCATION: Extracellular.
 CC -|- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -|- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L40156; AAA92021.1; -;
 CC HSP; P35247; I808.
 CC MGD; MGI:109515; Sftpd.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001304; lectin_c.
 CC Pfam; PF01391; Collagen; 3.
 CC Pfam; PF00059; lectin_c; 1.
 CC SMART; SM00034; CLECT; 1.
 CC PROSITE; PS00615; C-TYPE-LECTIN.1; 1.
 CC PROSITE; PS00041; C-TYPE-LECTIN.2; 1.
 CC GlycoProtDB; Calcium; Surface film; Gaseous exchange; Hydroxylation;
 CC Signal; Lectin; Collagen; Repeat; Coiled coil.
 CC SIGNAL 1 19
 CC BY SIMILARITY.
 CC CHAIN 20 374
 CC PULMONARY SURFACTANT-ASSOCIATED PROTEIN
 CC D.
 CC DOMAIN 45 221
 CC COLLAGEN-LIKE.
 CC FT DOMAIN 222 253
 CC COILED COIL (POTENTIAL).
 CC FT DOMAIN 278 374
 CC C-TYPE LECTIN (SHORT FORM).
 CC FT DISULFID 280 372
 CC BY SIMILARITY.
 CC FT DISULFID 350 364
 CC BY SIMILARITY.
 CC FT CARBOHYD 89 89
 CC N-LINKED (GLCNAC...) (POTENTIAL).
 CC SEQUENCE 374 AA; 37688 MW; FE034261263F43E4 CRC64;
 CC
 CC Query Match 22.9%; Score 340; DB 1; Length 374;
 CC Best Local Similarity 32.4%; Pred. No. 8.5e-22;
 CC Matches 80; Conservative 34; Mismatches 107; Indels 26; Gaps 6;
 CC
 CC QY 46 PGPAGDGEKGD---PGEGKHGKVGKMGKIGKELGMDGRNIGTKTPIGKKGDKGE 102
 CC DB 134 PGPAGGKPGVEGAPGQGSTGAKSTGPKGERGAPGVQAGPAGNAGAGPAGPQGA 193
 CC QY 103 KGLIGIPGKAGTVCDCGRYKRFVGLDI-SIARLKTSMKFKVKNVIAGI----- 152
 CC DB 194 PGRGPPGLKGRGVPGD---RKIGSGSLPDSALRQMEALKKGLQRLVAFSHYQK 249
 CC QY 153 -----RETEKFFIYVQEKYRSLTHCRIGMAMPKDEANLTIADYVAKSGFF 205
 CC DB 250 AALFPDGRSVGDKIFRTADSEKPFEDAEQMKQAGGOLASPRSATENAIOQLITAHN-K 308
 CC QY 206 RVFIVGNDLEREGQYMTDNTPLQNTSNNEGEPSDPYGHEDCVEMLSGGWNTTECHLT 265
 CC DB 309 RAFLSMIDVTGKTKTYTGTGPLY-YSNWPGEPPNNGAENCVEIFTGQWQNDKAGEQ 367
 CC QY 266 MYFVCEP 272
 CC DB 368 RLVICEF 374
 CC
 CC RESULT 4
 CC PSPD_HUMAN
 CC ID PSPD_HUMAN
 CC AC P35247; PRT; 375 AA.

DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
 GN SFTPD OR SFTP4 OR PSPD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=9315122; PubMed=8428971;
 RA Crouch E., Rust K., Veile R., Donis-Keller H., Grosso L.;
 RT "Genomic organization of human surfactant protein D (SP-D). SP-D is
 RT encoded on chromosome 10q22.2-23.1.";
 RL J. Biol. Chem. 268:2976-2983(1993).
 RN [2]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 214-243.
 RC TISSUE=Lung, and Amniotic fluid;
 RX MEDLINE=92322003; PubMed=1339284;
 RA Lu J., Willis A.C., Reid K.B.M.;
 RT "Purification, characterization and cDNA cloning of human lung
 RT surfactant protein D.";
 RL Biochem. J. 284:795-802(1992).
 RN [3]
 RN SEQUENCE OF 60-375 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Lung;
 RX MEDLINE=91378578; PubMed=1898081;
 RA Rust K., Grosso L., Zhang V., Chang D., Persson A., Longmore W.,
 RT Cai G.-Z., Crouch E.;
 RT "Human surfactant protein D: SP-D contains a C-type lectin
 RT carbohydrate recognition domain.";
 RL Arch. Biochem. Biophys. 290:116-126(1991).
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=99197291; PubMed=10368295;
 RA Hakansson K., Lim N.K., Hoppe H.-J., Reid K.B.M.;
 RT "Crystal structure of the trimeric alpha-helical coiled-coil and the
 RT three lectin domains of human lung surfactant protein D.";
 RL Structure 7:255-264(1999).
 CC -|- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
 CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
 CC -|- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -|- SUBCELLULAR LOCATION: Extracellular.
 CC -|- PTM: The N-terminus is blocked.
 CC -|- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -|- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L05485; AAB59450.1;
 CC EMBL; L05483; AAB59450.1; JOINED.
 CC EMBL; L05484; AAB59450.1; JOINED.
 CC EMBL; X65018; CAA46152.1;
 CC PIR; A45225; A45225.
 CC PIR; S18382; S18382.
 CC PDB; 1B08; 29-NOV-99.
 CC MIM; 178635;
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001304; lectin_c.

[illegible]

Query Match 22.1%; Score 328.5; DB 1; Length 371;
Best Local Similarity 29.5%; Pred. NO. 8e-21;
Matches 85; Conservative 35; Mismatches 95; Indels 73; Gaps 10;
QY 46 PGPKGNDGEKDP-----GEEKGKVGKVMGPKIGKELGMDGRNTGK-----TGP 93

Db 96 GPKGDTGPRGPGMGPGAGREGPSKSGMGPGPGTGGKGTGKGGVAGPQIGFPGP 155
QY 94 IGGKDGKGGKLLGIPGE-----KGKAG- 116
Db 156 SGLGKGGKAGPGTGGAGRGVTPSGAICPGPGSGARGPPGLKGRDGPGETGAKGESGL 215
QY 117 -----TVCDGGRYKRFVQGLDISIARLTKTSMKFKVKNVIAGIRETEKEKYIVQBE 166
Db 216 AEYNALQKRVILD-GHLRFQ-----AFQYKAVLFPDGOAVG-----EKIFKTAGAV 265
QY 167 KNTRESLTHORIGMLAMPKDEAANTLIADYVAKSGFFRVFVGNLDEREGYMTDNT 226
Db 266 KYSYDAEOLCREAGKGLASPRSAENAVTQW-RAQKNVLSNNDLSIEGRFYTPGE 324
QY 227 PLQVSNWNEGEP--SDPYGHEDCVEMLSGRWNDECHLTWYFVCEE 272
Db 325 ILV-YSNWADGEPNNSDEGQNCVEIFPDGKWNVDPCSKQLLVICF 371
RESULT 7
PSPA_RABIT STANDARD; PRT; 247 AA.
AC P12842;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPR).
GN SFTPA OR SFTPA OR SFTPI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88139348; PubMed=2830270;
RA Bogdaram V., Qing K., Mendelson C.R.;
RT "The major apoprotein of rabbit pulmonary surfactant. Elucidation of
RT primary sequence and cyclic AMP and developmental regulation.";
RL J. Biol. Chem. 263:2339-2347 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-NEW ZEALAND WHITE; TISSUE=Liver;
RC MEDLINE=92312742; PubMed=1616051;
RA Chen Q., Bogdaram V., Mendelson C.R.;
RT "Rabbit lung surfactant protein A gene: identification of a lung-
RT specific DNase I hypersensitive site.";
RL Am. J. Physiol. 262:L662-L671 (1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPA BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: J03542; AAA31465.1; -;
DR EMBL: L19387; AAA31468.1; -;
DR PIR: A29931; LNRBPS.

DR HSP; P35247; I808.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT A.
FT DOMAIN 27 99 COLLAGEN-LIKE.
FT DOMAIN 152 247 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 154 245 BY SIMILARITY.
FT DISULFID 223 237 BY SIMILARITY.
FT CARBOHYD 206 206 N-LINKED (GLCNAC...) (PROBABLE).
FT VARIANT 12 12 S -> P.
FT CONFLICT 57 60 GPMG -> APWA (IN REF. 2).
SQ SEQUENCE 247 AA; 289634054CB8CB4 CRC64;
Query Match 20.6%; Score 305.5; DB 1; Length 247;
Best Local Similarity 31.2%; Pred. No. 4.6e-19;
Matches 85; Conservative 30; Mismatches 104; Indels 53; Gaps 9;
QY 25 SGLGIDSDSPTAEVCATHI---SPGKGDGDEKGDPEEGKHGKVGKRGKIGKELGD 81
Db 5 SLATLISAPASDCTDYKDCIGSPGIPGTGSHGLPGRDGRGVKGGDPPGPMGPGG 64
QY 82 M-----GDRNICKTGPIGKKDKGKGLLIGIPGKAGKAGTVCDCGRYKRFVQGLD----- 132
Db 65 MPGLPGRDLGAPGVGERGDKE-----PGERGPPG-----LPAYLDELOA 108
QY 133 -----ISLARLTKSMKFKVKNVIAGIRETEKEKYIVQEKYRSLTHORIG 180
Db 109 TLIELRHRLQSIGVLSLQSGKAV-----GEKIFSTNGSQVNFDAIREVCARAG 158
QY 181 GLMLPKDEAANTLIADYVAKSGFFRVFVGNLDEREGYMTDNTPLQVSNWNEGEP 240
Db 159 GRIAPVSRLEENALASIVKERNY-AYLGAEGPTAGDYFYLDDPV-VYTNWYCPERP 216
QY 241 DPGHEDCVEMLSGRWNDECHLTWYFVCEE 272
Db 217 G-QGREKCVEMVYTDKWNKNCILQYLVICF 247
RESULT 8
MABA_RAT STANDARD; PRT; 238 AA.
AC P19999;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Mannose-binding protein A precursor (MBP-A) (Mannan-binding protein).
GN MBP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=86196130; PubMed=3009480;
RA Drickamer K., Dordal M.S., Reynolds L.;
RT "Mannose-binding proteins isolated from rat liver contain
RT carbohydrate-recognition domains linked to collagenous tails.
RT Complete primary structures and homology with pulmonary surfactant
RT apoprotein.";
RL J. Biol. Chem. 261:6878-6887 (1986).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;

RX MEDLINE=87137502; PubMed=3029088;
RA Drickamer K., McCreary V.;
RT "Exon structure of a mannose-binding protein gene reflects its
RT evolutionary relationship to the asialoglycoprotein receptor and
RT nonfibrillar collagens";
RL J. Biol. Chem. 262:2582-2589(1987).
RN [3]
RP SEQUENCE OF 18-42.
RX MEDLINE=87222358; PubMed=3584121;
RA Ikeda K., Sannoh T., Kawasaki N., Kawasaki T., Yamashina I.;
RT "Serum lectin with known structure activates complement through the
RT classical pathway";
RL J. Biol. Chem. 262:7451-7454(1987).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 124-238.
RX MEDLINE=92086855; PubMed=1721241;
RA Weis W.I., Kahn R., Fourme R., Drickamer K., Hendrickson W.A.;
RT "Structure of the calcium-dependent lectin domain from a rat mannose-
RT binding protein determined by MAD phasing.";
RL Science 254:1608-1615(1991).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 124-238.
RX MEDLINE=93063338; PubMed=1436090;
RA Weis W.I., Drickamer K., Hendrickson W.A.;
RT "Structure of a C-type mannose-binding protein complexed with an
RT oligosaccharide.";
RL Nature 360:127-134(1992).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 84-238.
RX MEDLINE=95219384; PubMed=7704532;
RA Weis W.I., Drickamer K.;
RT "Trimeric structure of a C-type mannose-binding protein.";
RL Structure 21:1227-1240(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 90-238.
RX MEDLINE=99119227; PubMed=9922165;
RA Ng K.K.-S., Park-Snyder S., Weis W.I.;
RT "Ca2+-dependent structural changes in C-type mannose-binding
RT proteins.";
RL Biochemistry 37:17965-17976(1998).
CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-
CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,
CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF
CC THE ANTIBODY.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL SPACE
CC OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M14105; AAA98781.1;
CC EMBL: M14104; AAA98781.1; JOINED.
CC FIR: B24791; LNR7MA.
CC DR PDB: 1NSB; 15-JAN-92.
CC DR PDB: 2MSB; 31-OCT-93.
CC DR PDB: 1AFA; 03-APR-96.
CC DR PDB: 1AFB; 03-APR-96.
CC DR PDB: 1AFD; 03-APR-96.
CC DR PDB: 1FTW; 07-FEB-95.
CC DR PDB: 1KMB; 12-FEB-97.
CC DR PDB: 2KMB; 12-FEB-97.
CC DR PDB: 3KMB; 12-FEB-97.

DR PDB: 4KMB; 12-FEB-97.
DR PDB: 1VTI; 10-JUN-96.
DR PDB: 1BCH; 17-JUN-98.
DR PDB: 1BCU; 17-JUN-98.
DR PDB: 1BUU; 09-SEP-98.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_2; 1.
KW Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;
KW Calcium; Collagen; Repeat; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 238
FT DOMAIN 39 88
FT DOMAIN 143 238
FT MOD_RES 43 43
FT MOD_RES 61 61
FT MOD_RES 67 67
FT MOD_RES 73 73
FT MOD_RES 78 78
FT DISULFID 145 234
FT DISULFID 212 226
FT CONFLICT 156 156
FT STRAND 127 136
FT STRAND 138 147
FT HELIX 148 149
FT STRAND 151 152
FT HELIX 158 168
FT STRAND 172 177
FT TURN 182 183
FT STRAND 186 187
FT TURN 188 189
FT STRAND 192 192
FT STRAND 198 198
FT TURN 200 201
FT TURN 207 208
FT STRAND 212 215
FT HELIX 217 219
FT STRAND 221 224
FT TURN 226 227
FT STRAND 230 237
SQ SEQUENCE 238 AA; 25308 MW; 1A927482B9A8CB3D CRC64;

Query Match 19.5%; Score 290; DB 1; Length 238;
Best Local Similarity 28.7%; Pred. No. 9.3e-18;
Matches 75; Conservative 49; Mismatches 105; Indels 32; Gaps 8;

QY 14 ILIVLFLIQIQLGLDIDSRPTAEVCATHTISPGKGDGDEKGDGEGKHGKVGKMGPK 73
DB 6 LLVLLCVSVSSSSQ--TCEETLKTC--VIAAGRGDRGCPKGEKGPQG-----GLR 55
QY 74 GKGELGDMGRGNIGTKTGPICKGDKGKGLLGPCKEKGAGTVCDCGYRKFVGOLDI 133
DB 56 GLQGPPLKGLPPGSGVAPSGQPGKQKGDG-----DSRAIEVKLANMEA 100
QY 134 SIARLTKSKFKVKNVITAGI--RETEKEFYIVOEKYNRESLTHCRTRGMLAMPKDEAA 191
DB 101 EINTLKSLEITNKLHAFSGKSKSKKFFVTNHPFFSKVKALCSELRTVAIPNAPE 160
QY 192 NTLIADYVAKSGFPRFVIGVNDLEREGYMTDNTPLQYNSWNEGEPSPYGHEDCYEM 251
DB 161 NKAIOE-VAKTS---AFLGITDEVTEGQFMVYVGGRL--TYSNKKKDEPDHGGGDCVTI 215
QY 252 LSSGRWNDETECHLTMYVVFCEP 272
DB 216 VDNGLWINDISQASHTAVCEP 236

RESULT 9
MABC_MOUSE

[3] SEQUENCE FROM N.A.
MEDLINE=90073571; PubMed=2590164;
Taylor M.E., Brickell P.M., Craig R.K., Summerfield J.A.;
"Structure and evolutionary origin of the gene encoding a human serum
mannose-binding protein.";
Biochem. J. 262:763-771(1989).
[4]
PARTIAL SEQUENCE.
TISSUE=Liver, and Plasma;
MEDLINE=95073978; PubMed=7982896;
Kurata H., Sannoh T., Kozutsumi Y., Yokota Y., Kawasaki T.;
"Structure and function of mannan-binding proteins isolated from
human liver and serum.";
J. Biochem. 115:1148-1154(1994).
[5]
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 108-248.
MEDLINE=95360730; PubMed=7634089;
Sheriff S., Chang C.Y., Ezekowitz R.A.;
"Human mannose-binding protein carbohydrate recognition domain
trimerizes through a triple alpha-helical coiled-coil.";
Nat. Struct. Biol. 1:789-794(1994).
[6]
VARIANT ASP-54.
MEDLINE=91269930; PubMed=1675710;
Sumiya M., Super M., Tabona P., Levinsky R.J., Arai T., Turner M.W.,
Summerfield J.A.;
"Molecular basis of opsonic defect in immunodeficient children.";
Lancet 337:1569-1570(1991).
[7]
VARIANTS ASP-54 AND GLO-57.
MEDLINE=93258313; PubMed=1304173;
Lipscombe R.J., Sumiya M., Hill A.V.S., Lau Y.L., Levinsky R.J.,
Summerfield J.A., Turner M.W.;
"High frequencies in African and non-African populations of
independent mutations in the mannose binding protein gene.";
Hum. Mol. Genet. 1:709-715(1992).
[8]
ERRATUM.
Lipscombe R.J., Sumiya M., Hill A.V.S., Lau Y.L., Levinsky R.J.,
Summerfield J.A., Turner M.W.;
Hum. Mol. Genet. 2:342-342(1993).
[9]
VARIANT ASP-54.
MEDLINE=93265124; PubMed=1303250;
Super M., Gillies S.D., Foley S., Sastry K., Schweinle J.E.,
Silverman V.J., Ezekowitz R.A.;
"Distinct and overlapping functions of allelic forms of human mannose
binding protein.";
Nat. Genet. 2:50-55(1992).
[10]
VARIANTS CYS-52; ASP-54 AND GLO-57.
MEDLINE=99374928; PubMed=10447262;
Gabolde M., Muralitharan S., Besmond C.;
"Genotyping of the three major allelic variants of the human
mannose-binding lectin gene by denaturing gradient gel
electrophoresis.";
Hum. Mutat. 14:80-83(1999).
-1- FUNCTION: BINDS MANNOSE AND N-ACETYLGALUCOSAMINE IN A CALCIUM-
DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,
BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF
THE ANTIBODY.
-1- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
-1- DISEASE: THERE IS AN ASSOCIATION BETWEEN LOW LEVELS OF MBL AND A
DEFECT OF OPSONIZATION WHICH RESULTS IN SUSCEPTIBILITY TO FREQUENT
AND CHRONIC INFECTIONS.
-1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
-1- CAUTION: THERE ARE EXTENSIVE DIFFERENCES BETWEEN THE REVISED
SEQUENCE IN REF.1 AND THAT PUBLISHED IN REF.2.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at EMBL.

14	QY	ILLVIFLQIQSLGLDIDNPFAVCATHISGPGKDDKEKDGPGEEKSHKGVGRMGPK	73
6	Db	LATFTLLTAVSLGECVD----KEVCLA---SSGIPCTPGSHGLPGRDGIRGKDGPP	57
74	QY	GIKGEIGDMGRGNIGTPIGKKGDKGEKGLLIGIPGKGAETVDCDGRVKFVGGQLDI	133
58	Db	GPMPGPPGMA--GPPGDQDMICAPGLGEERGEKGFGEPPG-----LPALHD--104	

[illegible]


```
FT DISULFID 39 39 INTERCHAIN (BY SIMILARITY).
FT DISULFID 156 247 BY SIMILARITY.
FT DISULFID 223 235 BY SIMILARITY.
SQ SEQUENCE 249 AA; 26471 MW; 29FC9F5927A6DD5 CRC64;

Query Match 19.24; Score 284.5; DB 1; Length 249;
Best Local Similarity 34.5%; Pred. No. 2.9e-17;
Matches 77; Conservative 31; Mismatches 94; Indels 21; Gaps 8;

QY 56 GDPEEKHGKVGMPGKNGELGDMGDRNGIKGTGPKCKGDKGKGLGIP---GEX 112
DQ 40 GPPGICNGIKGDRGDKGKGEKGP-----QGLRSGQGGPKMGPGTGPICGPIGQK 95
QY 113 KGAQVDCGRYKRVFGQIDISTARLKTSMKVN--VIAGIRETEKFFYIVVEKNYR 170
DQ 96 GDPE--NMGDVIKRLATS---ERATIQSELNQRNWLIFSLGKRVGKKAFFTKGKMPFN 150
QY 171 ESLTHCKIRGGLAMPKDEAANTLIADYVAKSGFFRVFVGNDLREGYMFNTDTEPL-Q 229
DQ 151 EVKTLCAQFGVRATPMNAENRALDLYTE----EAFILGIDQETGK--FVLDLTKGV 204
QY 230 NYSNWNQSPDPYGHEDCVDMSSGRWNDECHLTWYFVCEE 272
DQ 205 TYQWNGDPPNNSPGEHCVTLLSDGTWNDIACSGASFLTVCFF 247

RESULT 13
PSPA_HUMAN
ID PSPA_HUMAN STANDARD; PRT; 248 AA.
AC P07714;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP) (Alveolar proteinosis protein) (35 kDa pulmonary surfactant-
DE associated protein).
GN SFTPA1 OR SFTPA OR SFTP1 OR PSPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86250832; PubMed=3755136;
RA Flores J., Steinbrink R., Jacobs K., Phelps D., Kriz R., Reeny M.,
RA Sultzman L., Jones S., Teusch H.W., Frank H.A., Fritsch E.F.;
RT "Isolation and characterization of cDNA clones for the 35-kDa
RT pulmonary surfactant-associated protein."
RL J. Biol. Chem. 261:9029-9033(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014366; PubMed=2995821;
RA White R.T., Damm D., Miller J., Spratt K., Schilling J., Hawgood S.,
RA Benson B., Cordell B.;
RT "Isolation and characterization of the human pulmonary surfactant
RT apoprotein gene."
RL Nature 317:361-363(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92198680; PubMed=1372511;
RA Katyal S.L., Singh G., Locker J.L.;
RT "Characterization of a second human pulmonary surfactant-associated
RT protein SP-A gene."
RL Am. J. Respir. Cell Mol. Biol. 6:446-452(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
```

```
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13686; AAA60211.1; -
CC EMBL: K03475; AAA36520.1; -
CC EMBL: M30838; AAA36510.1; -
CC EMBL: M68519; AAA60319.1; -
CC PIR: A24622; LNHUP6.
CC PIR: A25720; LNHUP6.
CC PIR: B25720; LNHUP6.
CC HSSP: E19999; LYTI.
CC MIM: 178630; -
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001304; lectin_c.
CC Pfam: PF00059; lectin_c; 1.
CC SMART: SM00034; CLECT; 1.
CC PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
CC PROSITE: PS00641; C_TYPE_LECTIN_2; 1.
CC Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
CC Signal; Lactin; Collagen; Repeat; Polymorphism.
CC SIGNAL 1 20
CC CHAIN 21 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
CC DOMAIN 28 100 A.
CC C-TYPE LECTIN (SHORT FORM).
CC BY SIMILARITY.
CC BY SIMILARITY.
CC N-LINKED (GLCNAC. . .) (PROBABLE).
CC N -> T.
CC /FTID-VAR_004184.
CC L -> V.
CC /FTID-VAR_012231.
CC M -> T.
CC /FTID-VAR_004185.
CC D -> N.
CC /FTID-VAR_004186.
CC I -> V.
CC /FTID-VAR_004187.
CC C -> R.
CC /FTID-VAR_004188.
CC R -> W.
CC /FTID-VAR_012232.
CC Q -> K.
CC /FTID-VAR_012233.
CC A -> V (IN REF. 2).
CC CONFLICT 45 45 D -> H (IN REF. 2).
CC CONFLICT 54 54 P -> L (IN REF. 2).
CC CONFLICT 91 91 P -> A (IN REF. 3).
CC CONFLICT 100 100 P -> R (IN REF. 2; AAA36510).
CC CONFLICT 247 247 E -> D (IN REF. 3).
CC SEQUENCE 248 AA; 26214 MW; 6A9F0C3488BF3633 CRC64;

Query Match 19.1%; Score 284; DB 1; Length 248;
Best Local Similarity 29.9%; Pred. No. 3.2e-17;
Matches 78; Conservative 42; Mismatches 119; Indels 22; Gaps 9;

QY 16 LVFLQLQISGLDIDSRPTAFVCAHTTISPGKDGCKGKGDGKGVKRMGPKGI 75
DQ 6 LALNLLMAASGAACEVK---DVCVG---SPGPGTSGHGLPGDRGLKGDGPPPEP 59
QY 76 KGELGDMG-DRNGIKGTGPKCKGDKGKGLLIPGKAGTVCDCGRYKRVFGQIDIS 134
```

Db 60 MGPPGEMPCPPGNDGLPGAPGPGCEGK---GEPGERGPPGLPAHDE-----ELOAT 110
QY 135 IARUKTSKFKVNWIA---GIRETEKFFYIVQEEKYRSLTHCRIRGGMAMPKDEAA 191
Db 111 LHDFRHQILQTRGALSQSGTMTVGKVFSSQSGSITFDIAQACARAGGRIAVRNPEE 170
QY 192 NTLIADYVAKSGFFRVFVGNVLEREGGYMETDNTPLQNSNNEGSDPYGHEDQVEM 251
Db 171 NEATASFVKYNTY-AYVGLTEGSPGDFRYSQGTVPV-NVTNWRGEPAG-RGREQCQVEM 227
QY 252 ISSGRWNDECHLTWYFVCEP 272
Db 228 YTDGQWNRNCLYSRLICEF 248

RESULT 14
PSPA_CANFA
ID PSPA_CANFA STANDARD; PRT; 248 AA.
AC P06909;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSP-A).
GN SFTPAL OR SFTPA OR SFTPL.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND SIGNAL SEQUENCE CLEAVAGE SITE.
RX MEDLINE=86016705; PubMed=3863100;
RA Benson B., Haygood S., Schilling J., Clements J., Damm D., Cordell B.,
RA White R.T.;
RT "Structure of canine pulmonary surfactant apoprotein: cDNA and
RT complete amino acid sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 82:6379-6383(1985).
RN [2]
RP C-TYPE LECTIN DOMAIN.
RX MEDLINE=87115814; PubMed=3808053;
RA Parthy L.;
RT "Is lung surfactant protein a lectin-collagen hybrid?";
RL Nature 325:490-490(1987).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: M11769; AAA30887.1;
CC PIR: A25296; LNDGSP.
CC HSSP: P35247; 1B08.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001304; Lectin_C.
CC Pfam: PF01391; Collagen; 1.
CC Pfam: PF00059; Lectin_C; 1.
CC SMART: SM00034; CLECT; 1.
CC PROSITE: PS00615; C-TYPE LECTIN_1; 1.

DR PROSITE: PS50041; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Caseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat.
FT SIGNAL 1 17
FT CHAIN 18 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT A.
FT DOMAIN 28 100 COLLAGEN-LIKE.
FT C-TYPE LECTIN (SHORT FORM).
FT DISULFID 153 248 BY SIMILARITY.
FT DISULFID 155 246 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT CARBOHYD 20 20 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC... (PROBABLE).
SQ SEQUENCE 248 AA; 26268 MW; 340FE95D4E2502C0 CRC64;

Query Match 18.9%; Score 280; DB 1; Length 248;
Best Local Similarity 30.0%; Pred. No. 6.9e-17;
Matches 81; Conservative 36; Mismatches 115; Indels 38; Gaps 11;
QY 14 ILLVFLIQTSQGLDIDSRPTAEVCATHTISPGKGGDDGEGKGGKGVKGMGP 73
Db 6 LALATLLMVSGI-----ENNTKDVGVG---NPGIPGTPGSHGLPGRDGRGVKGP 57
QY 74 GIKGELGDM-GDRGNIGKTGPIGKKGDKGKGLLGPGEKAGTVCDCGRYKFKVG 132
Db 58 GLPGPPGMPGHPGPMGTGAPVAGERGEK---GEPGERGPPG-----LPASLD 104
QY 133 ISIALKTSMKFKVKNVI---AGIRETEE-----KFYIVQEEKYRSLTHCRIRGM 182
Db 105 ---EELQTLHLRHQILQTMGVLSHESLLVGRKVFSSNAQSFINDIQELCAGAGG 161
QY 183 LAMPKDEANLTIADYVAKSGFFRVFVGNVLEREGGYMETDNTPLQNSNNEGSDP 242
Db 162 IAAPMSPEENEAIVSVKVKYNTY-AYLGLVSPDSGDFQYMDGAPV-NVTNYPGEPRG- 218
QY 243 YGHEQCEVEMLSGRWNDECHLTWYFVCEP 272
Db 219 RKEQCEVEMYTDGQWNRNCLYSRLICEF 248

RESULT 15
MABC_RAT
ID MABC_RAT STANDARD; PRT; 244 AA.
AC P08661;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)
DE (RA-reactive factor P28A subunit) (RAFF/P28A).
GN MBL2
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=92299655; PubMed=1607365;
RA Wada M., Itoh N., Ohta M., Kawasaki T.;
RT "Characterization of rat liver mannan-binding protein gene."
RL J. Biochem. 111:66-73(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=86196130; PubMed=3009480;
RA Drickamer K., Dordal M.S., Reynolds L.;
RT "Mannose-binding proteins isolated from rat liver contain
RT carbohydrate-recognition domains linked to collagenous tails.
RT Complete primary structures and homology with pulmonary surfactant
RT apoprotein."
RL J. Biol. Chem. 261:6878-6887(1986).
RN [3]
RP SEQUENCE FROM N.A.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:31:36 ; Search time 20.65 Seconds
(without alignments) 327.646 Million cell up

Title: US-09-600-932-2
Perfect score: 1484
Sequence: 1 MNGFASILLRRNQVILLVFL.....NDTECHLTWYFVCEFKKKK 277

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Issued_Patents_AA:*
1: /cn2.6/ptdata/2/iaa/5A_COMB.pcp.*
2: /cn2.6/ptdata/2/iaa/5B_COMB.pcp.*
3: /cn2.6/ptdata/2/iaa/6A_COMB.pcp.*
4: /cn2.6/ptdata/2/iaa/6B_COMB.pcp.*
5: /cn2.6/ptdata/2/iaa/PCUTS_COMB.pcp.*
6: /cn2.6/ptdata/2/iaa/backfiles1.pcp.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	328.5	22.1	351	3	US-09-011-735-1	Sequence 1, Appl
2	289.5	19.5	248	4	US-09-196-603C-2	Sequence 2, Appl
3	191.5	12.9	1024	3	US-08-931-820-2	Sequence 2, Appl
4	191.5	12.9	1366	3	US-08-963-825-19	Sequence 19, Appl
5	191.5	12.9	1366	4	US-09-570-573-19	Sequence 19, Appl
6	191.5	12.9	1366	4	US-08-548-608-19	Sequence 19, Appl
7	187.5	12.6	489	2	US-08-794-795-7	Sequence 7, Appl
8	187.5	12.6	489	4	US-09-249-200-7	Sequence 7, Appl
9	187.5	12.6	518	1	US-08-332-367B-2	Sequence 2, Appl
10	187.5	12.6	518	3	US-08-893-467A-2	Sequence 2, Appl
11	187	12.6	161	3	US-09-011-735-6	Sequence 6, Appl
12	181.5	12.2	128	4	US-09-227-357-190	Sequence 190, App
13	180	12.1	107	6	5514582-17	Patent No. 5514582
14	180	12.1	557	3	US-09-320-095-10	Sequence 10, Appl
15	180	12.1	557	4	US-09-523-487-10	Sequence 10, Appl
16	179	12.1	495	2	US-08-734-795-2	Sequence 2, Appl
17	179	12.1	495	4	US-09-249-200-2	Sequence 2, Appl
18	179	12.1	520	2	US-08-794-795-6	Sequence 6, Appl
19	179	12.1	520	4	US-09-249-200-6	Sequence 6, Appl
20	177.5	12.0	1060	3	US-08-931-820-3	Sequence 3, Appl
21	177.5	12.0	1418	3	US-08-963-825-20	Sequence 20, Appl
22	177.5	12.0	1418	4	US-09-010-999-1	Sequence 1, Appl
23	177.5	12.0	1418	4	US-09-570-573-20	Sequence 20, Appl
24	177.5	12.0	1418	4	US-09-548-608-20	Sequence 20, Appl
25	177	11.9	294	4	US-09-188-930-294	Sequence 294, App
26	176.5	11.9	532	1	US-08-494-168-9	Sequence 9, Appl
27	176.5	11.9	595	4	US-09-219-849-48	Sequence 48, Appl

ALIGNMENTS

```

RESULT      1
US-09-011-735-1
; Sequence 1, Application US/09011735B
; Patent No. 6110708
; GENERAL INFORMATION:
;   APPLICANT: Wakamiya, No. 6110708uraka
;   TITLE OF INVENTION: Recombinant Conglutinin and Producing Method thereof
;   FILE REFERENCE: 19036/34348
;   CURRENT APPLICATION NUMBER: US/09/011.735B
;   CURRENT FILING DATE: 1998-05-22
;   EARLIER APPLICATION NUMBER: JP 7-209698
;   EARLIER FILING DATE: 1995-08-17
;   NUMBER OF SEQ ID NOS: 6
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 351
;   TYPE: PRT
;   ORGANISM: Bovine
US-09-011-735-1

```

Query Match	22.1%	Score 328.5;	DB 3;	Length 351;
Best Local Similarity	29.2%;	Prod. No. 4.1e-25;		
Matches	84;	Conservative 33;	Mismatches 98;	Indels 73; Gaps 9;
QY	46	PGPKGDDEKGD	-----GEGKHGKVGWMPKGIKIGELGMDGRNICK-----	TGP 93
Db	76	PGPKGTPRGPP	CMPCAGPAGREPSGKSGMPPGTPGPKETGPKGGVGARGIQGPPG	135
QY	94	IGKKGDKEGKLLG	IPCEKKGAG-----	116
Db	136	SLGLKEKAGP	ETCAPRGAVGTPGSAIGQGPSPGARGPPLKCDKGDPCGTGAGSEGL	195
QY	117	-----	TVDCGRYKRFQGLD	ISIARLKTSMTKVNKIAGITEERKFYIVQEE 166
Db	196	AEYNALKQRV	IILD-GHLRRFON-----AFSQYKAVLPDPQAVG-----	EKIFKTAGAV 245
QY	167	KNYRESLTHCR	IRGGMLAMPKDEAANTLIADYVAKSGFRFVFGVNDLGRGQWFMFDNT	226
Db	246	KVSYDAQLCREAK	QOLASPRSAENAVQWV-RAQEKNAIYSMDISTEGRTYPTGE	304
QY	227	PLQYNNNBE	P--SDPYGHEDCVMELSSGRWNDDTECHLTWVFCF	272
Db	305	ILV-YENWADGEP	NNDDGQPCNCEVIFPDGKWNQVPGSKOLLVICEF	351

RESULT 2
US-09-198-603C-2
; Sequence 2, Application US/09198603C
; Patent No. 6337193

; GENERAL INFORMATION:
; APPLICANT: TOLLY, Raymond E.
; APPLICANT: CALTAGIRONE, G. Thomas
; APPLICANT: MOYER, Shawn S.
; APPLICANT: RONNING, Michael I.
; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: A7290
; CURRENT APPLICATION NUMBER: US/09/198,603C
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Human
US-09-198-603C-2

Query Match 19.5%; Score 289.5; DB 4; Length 248;
Best Local Similarity 31.1%; Pred. No. 2.2e-21;
Matches 75; Conservative 37; Mismatches 100; Indels 29; Gaps 7;
QY 34 PTAEEVCAHTIS--PGPKGDDGKGGEGGKHGKVGMRGPKGIGELGMDGRGNIGKT 91
DB 33 PAVIACSSPGINGFGKRGDRGTHGKEGPGQ-----GLRGLQPPKGLGPPGPGFS 85
QY 92 GPIKKKDKGKGLGIPGKKGAGTVCDCQRYKRVGQLDISTARLKTSMKFKVKNVIAG 151
DB 86 GSPGPKGQKGDG-----KSPDGSSLAASERK---ALQTEMAFIKKWLTESLG---- 131
QY 152 IRETEKEFYIVQEEKYKRESLTHCRIRGGLAMPKDEANLTIADYVAKSGFRFVIGV 211
DB 132 -KQGNKFLTNGELTMEFKVXALCVKFOASVATPRNAENGAIQNLKE----EAFLGI 186
QY 212 NDLEREQYMFDTNPIONTSNNNEGSDPDYGHEDCVMLSSGRWNDECHLMTYFVCE 271
DB 187 TDKTESQFVDITGNRL--TYNNNEGEPNAGSDDCVLLKNGQWNPVCSITSHLAVE 245
QY 272 F 272
DB 246 F 246

RESULT 3
US-09-931-820-2
; Sequence 2, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-2
Query Match 12.9%; Score 191.5; DB 3; Length 1024;
Best Local Similarity 43.5%; Pred. No. 1.1e-10;
Matches 40; Conservative 9; Mismatches 22; Indels 21; Gaps 2;
QY 46 PGPKGDDGKGGEGGKHGKVGMRGPKGIGELGMDGRGNIGKTGPIGKK----- 97
DB 479 PGPSPAGEVGKZGKGLHGEGLPGAGRGGRGPGGESGAAGTGPICSRGSPGPGP 538
QY 98 -GDKGKGLL-----GIPGKKGKAG 116
DB 539 DGNKGEPGVGAVGTAGPSGSGLPGERGAAG 570
RESULT 4
US-08-963-825-19
; Sequence 19, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adga C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: collagen alpha 2- type I
US-08-963-825-19
Query Match 12.9%; Score 191.5; DB 3; Length 1366;
Best Local Similarity 43.5%; Pred. No. 1.6e-10;
Matches 40; Conservative 9; Mismatches 22; Indels 21; Gaps 2;

```

Query Match      12.9%; Score 191.5; DB 4; Length 1366;
Best Local Similarity 43.5%; Pred. No. 1.6e-10;
Matches 40; Conservative 9; Mismatches 22; Indels 21; Gaps

QY 46 PGKAGDGGKGPDPGEGKHKHVGVMOPKIKETGLDMCDRNGTGTGPIGK----- 97
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 PGSGPAGEVGKPGERGLHGEFGLPGPAGPRERGGPGESGAAGTGTGSRGSPGPG 617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 98 -GDKGEKGLL-----GIPCEKAG 116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 DNKGKPGVGVAGTGAAGSPGSPGLPGERGAAG 649

```

RESULT 7
US-08-794-795-7
; Sequence 7, Application US/08794795
; Patent No. 5916766
; GENERAL INFORMATION:
; APPLICANT: Eishourlagy, Nabil

```
; APPLICANT: Adamou, John
; APPLICANT: Gross, Mitchell
; APPLICANT: Lyoko, Paul
; TITLE OF INVENTION: Human Macro Scavenger Rec
; TITLE OF INVENTION: Human Macro Scavenger Rec
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/794,795
; FILING DATE: 04-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: ATG50009P
; FILING DATE: 22-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50009
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-794-795-7

Query Match 12.6%; Score 187.5; DB 2; Length 489;
Best Local Similarity 40.8%; Pred. No. 9.9e-11;
Matches 40; Conservative 8; Mismatches 29; Indels 21; Gaps 1;

QY 46 PGPKGDDGKGGDPGEGKHGKVGKGMGPKGKIGEL-----GDMGD 84
Db 141 PGPAEKGKGAAGRDGTPGVGPGPGSKGAGLQGLTGAPGKGATGAPGPRGKGS 200

QY 85 RGNIGKTGPKGKGDKGKGLGIPGKGGKAGTVDCG 122
Db 201 KGDIGLTGPKGEGHGTGDKGDLGLPGNKGDMGKMGDTG 238

RESULT 8
US-09-249-200-7
; Sequence 7, Application US/09249200
; Patent No. 5197931
; GENERAL INFORMATION:
; APPLICANT: ELISHOURBAGY, NABIL
; APPLICANT: ADAMOU, JOHN
; APPLICANT: GROSS, MITCHELL
; APPLICANT: LYOKO, PAUL
; TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
```

```
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,200
; FILING DATE: 12-FEB-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/794,795
; FILING DATE: 04-FEB-1997
; APPLICATION NUMBER: 60/017,699
; FILING DATE: 23-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: ATG-50009-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-249-200-7

Query Match 12.6%; Score 187.5; DB 4; Length 489;
Best Local Similarity 40.8%; Pred. No. 9.9e-11;
Matches 40; Conservative 8; Mismatches 29; Indels 21; Gaps 1;

QY 46 PGPKGDDGKGGDPGEGKHGKVGKGMGPKGKIGEL-----GDMGD 84
Db 141 PGPAEKGKGAAGRDGTPGVGPGPGSKGAGLQGLTGAPGKGATGAPGPRGKGS 200

QY 85 RGNIGKTGPKGKGDKGKGLGIPGKGGKAGTVDCG 122
Db 201 KGDIGLTGPKGEGHGTGDKGDLGLPGNKGDMGKMGDTG 238

RESULT 9
US-08-392-367B-2
; Sequence 2, Application US/08392367B
; Patent No. 5691197
; GENERAL INFORMATION:
; APPLICANT: TRYGGVASON, KARL
; APPLICANT: ELICMAA, OUTI
; APPLICANT: KAANGAS, MAARIT
; TITLE OF INVENTION: An isolated DNA Sequence For a
; Patent No. 5691197
; TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
; TITLE OF INVENTION: a Collagenous Domain and the
; TITLE OF INVENTION: Polypeptide Chain Encoded by
; TITLE OF INVENTION: such a Sequence
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 720 Kb storable
```


7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:21 ; Search time 42.99 Seconds
(without alignments)
1114.668 Million cell updates/sec

Title: US-09-600-932-2

Perfect score: 1484

Sequence: 1 MNGFASLLRNQFILLVLFL.....NUTECHLIMFYVCEFTKKK 277

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1484	100.0	277	4 Q9Y627	Q9Y627 homo sapien
2	728	49.1	272	11 Q9DC75	Q9DC75 mus musculus
3	716.5	48.3	271	4 Q9BWP8	Q9BWP8 homo sapien
4	338.5	22.8	378	6 Q9NIX4	Q9NIX4 sus scrofa
5	337	22.7	254	13 Q9BTA4	Q9BTA4 gallus gall
6	334.5	22.5	251	13 Q9I908	Q9I908 brachydanio
7	327.5	22.1	238	13 Q57451	Q57451 gallus gall
8	325	21.9	335	6 Q97748	Q97748 bos taurus
9	316	21.3	256	13 Q9I909	Q9I909 cyprinus ca
10	307	20.7	246	13 Q9I907	Q9I907 carassius a
11	297.5	20.0	236	6 Q28518	Q28518 macaca mula
12	296	19.9	248	6 Q9TT06	Q9TT06 ovis aries
13	294	19.8	248	6 Q9N0G1	Q9N0G1 equus cabal
14	291	19.6	245	6 Q28517	Q28517 macaca mula
15	289.5	19.5	248	4 Q96KE4	Q96KE4 homo sapien
16	289	19.5	248	6 Q95L88	Q95L88 equus cabal

17	288.5	19.4	248	4 Q96TF7	Q96TF7 homo sapien
18	282.5	19.0	224	11 Q9Z294	Q9Z294 rattus sp.
19	282.5	19.0	248	4 Q96TF8	Q96TF8 homo sapien
20	281.5	19.0	248	4 Q96TF9	Q96TF9 homo sapien
21	273	18.4	742	4 Q9BYH7	Q9BYH7 homo sapien
22	272.5	18.4	248	11 Q9C011	Q9C011 mus musculus
23	256	17.3	240	6 Q9XSK3	Q9XSK3 sus scrofa
24	211	14.2	268	2 Q9F685	Q9F685 streptococc
25	207	13.9	689	4 Q14055	Q14055 homo sapien
26	206.5	13.9	222	13 Q90XB2	Q90XB2 gallus gall
27	205	13.8	645	9 Q9XKB9	Q9XKB9 escherichia
28	203	13.7	688	11 Q07643	Q07643 bacterioph
29	203	13.7	688	11 Q07643	Q07643 mus musculus
30	203	13.7	1723	5 Q9GQ31	Q9GQ31 hydra atten
31	198.5	13.4	479	2 Q9LA62	Q9LA62 escherichia
32	196.5	13.2	289	4 Q9HXJ2	Q9HXJ2 homo sapien
33	196	13.2	1779	5 Q9VMV4	Q9VMV4 streptococc
34	195.5	13.2	465	2 Q9AGC4	Q9AGC4 drosophila
35	193.5	13.0	291	2 Q9F690	Q9F690 streptococc
36	193	13.0	1549	11 Q60444	Q60444 cricetus
37	191.5	12.9	1366	4 Q15177	Q15177 homo sapien
38	190.5	12.8	410	2 Q9F691	Q9F691 streptococc
39	190	12.8	380	2 Q9F692	Q9F692 streptococc
40	189.5	12.8	358	5 Q23222	Q23222 caenorhabdi
41	189	12.7	397	2 Q9F686	Q9F686 streptococc
42	189	12.7	404	6 Q95J96	Q95J96 macaca mula
43	189	12.7	2944	11 Q63870	Q63870 mus musculus
44	188	12.7	291	5 Q23422	Q23422 caenorhabdi
45	188	12.7	310	13 Q90612	Q90612 gallus gall

ALIGNMENTS

RESULT 1
Q9Y627 PRELIMINARY; PRT; 277 AA.
ID Q9Y627
AC Q9Y627
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLLECTIN 34.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9240769; PubMed=10224141;
RA Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Yamazaki H.,
RA Keshi H., Sakai Y., Fukuchi A., Sakamoto T., Wakamiya N.;
RT "Molecular cloning of a novel collectin from liver (CL-L1).";
RL J. Biol. Chem. 274:13681-13689(1999).
DR EMBL; AB002631; BAA81747.1; -;
DR HSSP; P20693; IHLJ.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 1.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 277 AA; 30733 MW; 9736661CEBDC5C25 CRC64;

Query Match 100.0%; Score 1484; DB 4; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.6e-128;
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNGFASLLRNQFILLVLFLVLQIQLGIDSRPTAEVCATHTISPGKGDGEGKDGGE 60
Db 1 MNGFASLLRNQFILLVLFLVLQIQLGIDSRPTAEVCATHTISPGKGDGEGKDGGE 60


```
RESULT 4
Q9N1X4 AC Q9N1X4 PRELIMINARY; PRT; 378 AA.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN D PRECURSOR.
GN SFTPD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP MEDLINE=20109098; PubMed=10640760;
RA van Eijk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.*;
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF132496; AAF22145.2; -.
DR HSP: P35247; I908.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 378 LUNG SURFACTANT PROTEIN D.
SQ SEQUENCE 378 AA; 37986 MW; 3504E8C1E56C341D CRC64;

Query Match 22.8%; Score 338.5; DB 6; Length 378;
Best Local Similarity 32.8%; Pred. No. 5.7e-23;
Matches 81; Conservative 35; Mismatches 110; Indels 21; Gaps 6;

QY 45 SPGKGGDGPKGDPGEKHKVGMGPKIGKELGDMGR---GNIGTKTPIGKKGDKG 101
DB 134 TPFGPGTGPGEVSGALMGQSTGARGPAGLKGKGPAGSGAGAPAGATGPGQ 193
QY 102 EKGLLIPGEGKAGTVCDCG-----RYKFGQLDISIARLKTSMKFKVNV-- 148
DB 194 PSSARGPGLKGRGPPGERGAKGESLPGIARLQGVETLQGVQLQKAFSQYKVEL 253
QY 149 IAGIRETEEFYIVOEKNYRESLTHCRIRGMLAMPKDEAANTLIADYVAKSGFFRVF 208
DB 254 PFNGRGVGEKIFKGTGFEQDAQVCTQAGGQNASPRSETNEALSQLVTAQN-KA 312
QY 209 IGVDLEREGQYFTDNTPLQYNNWNEGEPSD---PYGHEDCVEMLSGRWNTDTECHLT 265
DB 313 LSWTDIKTEGNFYPTGPELV-YANWAPGEPNNGSSGAENCVEIFPNKGKWDKACGEL 371
QY 266 MYFVCEP 272
DB 372 RVLICEF 378

RESULT 5
Q98TA4 AC Q98TA4 PRELIMINARY; PRT; 254 AA.
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE MANNULOSE-BINDING LECTIN PROTEIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;

RESULT 6
Q91908 AC Q91908 PRELIMINARY; PRT; 251 AA.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MANNULOSE-BINDING-LIKE LECTIN PRECURSOR.
GN MBL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjoldt K.;
RA "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.*;
RT structure predicts affinity for galactose.*;
```

```
OC Gallus
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjoldt K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.*;
RL Immunogenetics 51:955-964(2000).
DR EMBL: AF231714; AAK30298.1; -.
DR HSP: P19999; IYTT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000005; HTHARAC.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
DR PROSITE: PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
KW Signal; Lectin.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 254 MANNULOSE-BINDING LECTIN PROTEIN.
FT VARIANT 234 234 D->V.
SQ SEQUENCE 254 AA; 27376 MW; C924428643441AED CRC64;

Query Match 22.7%; Score 337; DB 13; Length 254;
Best Local Similarity 31.8%; Pred. No. 4.7e-23;
Matches 92; Conservative 54; Mismatches 83; Indels 60; Gaps 15;

QY 6 SLLRRNQFTLLVFLQLQSLGIDISDRPTAEV-----CATHIS--PGPKDD---G 53
DB 2 TLQPSALLCLSLMATSL-LTTD-KPEKMYSCPIIQCSAPAVNGLPGDRGDPKG 59
QY 54 EKDPGE--EGKHGKVGKMGPKIGKELGDMGRGNIGTKTPIGKKGDKGKGL----- 106
DB 60 EKDPGEGLRGLOGLPKAGPKAGKGEV-----GPOGEGKGGKGGVIVTDDL 107
QY 107 --GIPGKAGTVCDCGGRYKFGVQLDISIARLKTSMKFKVNVNVIAGIRETEEFYIV 163
DB 108 HRQITDEAKIRVLEDDLRYKKAL-----SLKDVNV-----GKKMFVST 148
QY 164 QEKNYRESLTHCRIRGMLAMPKDEAANTLIADYVAKSGFFRVFVGVNLEREGQYMT 223
DB 149 GKYNFEKSKLAKAGSVLASPRNEAENTALKOLIDPSS--QAYIGISDAQTEGFMYL 206
QY 224 DNTPLQYNNWNEGEPSDYPGHEDCVEMLSGRWNTDTECHLTMTVFCVCE 271
DB 207 SGGPL-TYSNWKPGEPNN-HKNECAVIEDSGKNDLDCSNISNIFICE 253

RESULT 6
Q91908 AC Q91908 PRELIMINARY; PRT; 251 AA.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MANNULOSE-BINDING-LIKE LECTIN PRECURSOR.
GN MBL.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjoldt K.;
RA "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.*;
```

```

RL Immunogenetics 51:955-964(2000).
DR EMBL: AF227736; AAF63469.1; -.
DR HSP; P19999; IAFB.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 1.
DR DR: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS0041; C_TYPE_LECTIN_2; 1.
KW Signal; Lectin.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 251 MANNOSE BINDING-LIKE LECTIN.
FT VARIANT 21 21 M -> L.
FT VARIANT 204 204 K -> N.
SQ SEQUENCE 251 AA; 26829 MW; 12D0ABD06B6E3B11 CRC64;

Query Match 22.5%; Score 334.5; DB 13; Length 251;
Best Local Similarity 30.4%; Pred. No. 7.9e-23;
Matches 79; Conservative 49; Mismatches 107; Indels 25; Gaps 8;

QY 14 ILLVFLIQISGLDIDSRPTAEVCATHTISPGPKGDDGKGPGEKGKGVGRGPK 73
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 74 GIKGELGMDGRNIGKGTGKKGKGLGIPGKAGVCDGGRYKRVGQLDI 133
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 134 STARLTKSMKFKVNI--AGIRETEKFFYIVQEEKNYRESLTHCRIRGMLAMPKDEAA 191
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 110 ELQKLSKIALIEKVVKTKYKGVQYVTDVEETDKGMGYCSNGGALVLPRTLEE 169
QY 192 NTLIADYVAKSGFRVFGVANDLEREGYMTDNTPLQNYSNNEGEPSDPYGHEDCVEM 251
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 170 NALLKVFVS-SAFKRLFRITRDEKEGEFVDIRKKL-TFTNGPQPNQYKAGDQCGAI 227
QY 252 LSSGRWNTDCHLTMYFVCE 271
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 228 ADSGLMDVDSOSLYPIICE 247
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 7
Q57451 PRELIMINARY; PRT; 238 AA.
AC Q57451;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE MANNAN-BINDING LECTIN (FRAGMENT).
GN CMBL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=LIVER;
RA Laursen S.B.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF022226; BAB94071.1; -.
DR HSP; P19999; 1YTT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS0041; C_TYPE_LECTIN_2; 1.
KW Lectin.
FT NON_TER 238 238
SQ SEQUENCE 238 AA; 25645 MW; E5C9B5197AAE64E3 CRC64;

```

```

Query Match 22.1%; Score 327.5; DB 13; Length 238;
Best Local Similarity 32.1%; Pred. No. 3.3e-22;
Matches 80; Conservative 48; Mismatches 70; Indels 51; Gaps 12;

QY 39 CAHTIS--PGFKGDD--GEKGPGE--EGKHGKVGKMGPKGKIGELGDMGDRGNIGKT 91
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 24 CSAPAVNGLPGDRGDKGKGPGEGLRGLQGLPGKAGPQLKGEV----- 71
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 92 GPTGKKGDKGKGLL-----GIPGKAGKAGTV-CDGGRYKRVGQLDISIRLKTSMK 143
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 72 GPQGEKGQGERGIVVTDLHRQITDLKAKIRVLEDDLSRYKKAL-----SLK 119
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 144 EVKNVIAGIRETEKFFYIVQEEKNYRESLTHCRIRGMLAMPKDEAAATLIADYVAKSG 203
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 120 DVNI-----GKMFVSTGKKYNEFKGSLCAKAGSVLASPRNEAENTALKDLIDPSS 172
QY 204 FFRVFGVANDLEREGYMTDNTPLQNYSNNEGEPSDPYGHEDCVEMLSGRWNTDEC- 262
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 173 --QAYGISDAQTEGREFMYLSGGPL-TYSNKKPQEPNN-HKNECAVIEDSGKWNOLDCS 228
QY 263 HLTMFVCE 271
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 229 NSNIFICE 237
DB :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 8
Q57748 PRELIMINARY; PRT; 335 AA.
AC Q57748;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONGLUTININ.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
PT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93277452; PubMed=7684896;
RA Lu J., laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94128104; PubMed=8297370;
RA Kawasaki N., Itoh N., Kawasaki T.;
RT "Gene organization and 5'-flanking region sequence of conglutinin: a
C-type mammalian lectin containing a collagen-like domain.";
RL Biochem. Biophys. Res. Commun. 198:597-604(1994).
DR EMBL: D25302; BAA04983.1; -.
DR EMBL: D25296; BAA04983.1; JOINED.
DR EMBL: D25297; BAA04983.1; JOINED.
DR EMBL: D25299; BAA04983.1; JOINED.
DR EMBL: D25300; BAA04983.1; JOINED.
DR EMBL: D25301; BAA04983.1; JOINED.
DR HSP; P35247; IB08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 2.

```

```

DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS0041; C_TYPE_LECTIN_2; 1.
FT CONFLICT 236 E -> V (IN REF. 2).
SQ SEQUENCE 335 AA; 34702 MW; 39DA3A0BC76C134C CRC64;

Query Match 21.9%; Score 325; DB 6; Length 335;
Best Local Similarity 33.8%; Pred. No. 8.6e-22;
Matches 80; Conservative 33; Mismatches 96; Indels 23; Gaps 8;

QY 46 PGPFGDDEKGDGEGSKGKVGKMGPKGKIGKELGDMGDRGNIGKTPTGKKGDKGKGL 105
Db 117 PGPGLKGEKAGPAGAGVTPGSA-----IGPQPGSARGPGLAGDRDGPGE 170
QY 106 LGIPGKAGKAG-----TVCDGGRYKFKVQGLDISIARLKTSMKFKVNIAGIRETEE 157
Db 171 TGAKGESGLAEVNAKQRTILD-CHLRRFQN-----AFSQYKKAIVFPDQAVG-----E 220
QY 158 KYYIVQEEKNYRESLTHCRIGMLAMPKDEAANTLIADYVAKSGFRFVIGVNDLRE 217
Db 221 KIFKTAGAVKSYDAEQLCREAKQLASPRSSAENEAQVQMV-RAQEKNAIYLSMNDISTE 279
QY 218 GQYMTDNTPLQYNNWNGEP--SDPYGHEDCVEMLSGGRWNTTECHLTMFVCEFF 272
Db 280 GRFTYPTGEILV-YSNWADGEPNNSDEGPENCVEIFPDGKNDVPCSKQLLVICF 335

RESULT 9
Q91909
ID Q91909 PRELIMINARY; PRT; 256 AA.
AC Q91909;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MANNOSE BINDING-LIKE LECTIN PRECURSOR.
GN MBL.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjold K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.";
RL Immunogenetics 51:955-964(2000).
DR EMBL; AF227737; AAF63468.1; -.
DR HSSP; P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS0041; C_TYPE_LECTIN_2; 1.
KW Signal; Lectin.
FT SIGNAL 1 23 POTENTIAL.
FT VARIANT 24 256 MANNOSE BINDING-LIKE LECTIN.
FT CHAIN 235 235 S -> T.
SQ SEQUENCE 256 AA; 26934 MW; D019291D1167730D CRC64;

Query Match 21.3%; Score 316; DB 13; Length 256;
Best Local Similarity 31.8%; Pred. No. 4.1e-21;
Matches 84; Conservative 39; Mismatches 119; Indels 24; Gaps 8;

QY 14 ILLVFLLIQISGLIDISDPTAEVCATHTISPGKGDGEGKKGKVGKMGPK 73
Db 1 LLLVFLLIQISGLIDISDPTAEVCATHTISPGKGDGEGKKGKVGKMGPK 52
QY 74 GIKGELGDMGDRGNIGKTPTGKKGDKGKGLLIGIPGKAGTVCDCGRYKFKVGLDI 133
Db 53 GEKGESG-VSVQPGPKAGPAGTAGEKGERGSGGQSGPGS-----ESVLESLS 101
QY 134 STARLKTSMKFKVKNV--IAGIRETEEFYIVQEEKNYRESLTHCRIGMLAMPKDEA 191
Db 102 ETQQLAKIATEFEKVSCHFRKVGQYITDGVVGNFDGLKSCMEFGSTWVSPRTSAE 161

Db 11 LLLVFLLIQISGLIDISDPTAEVCATHTISPGKGDGEGKKGKVGKMGPK 62
QY 74 GIKGELGDMGDRGNIGKTPTGKKGDKGKGLLIGIPGKAGTVCDCGRYKFKVGLDI 133
Db 63 GEKGESG-VSVQPGPKAGPAGTAGEKGERGSGGQSGPGS-----ESVLESLS 111
QY 134 STARLKTSMKFKVKNV--IAGIRETEEFYIVQEEKNYRESLTHCRIGMLAMPKDEA 191
Db 112 ETQQLAKIATEFEKVSCHFRKVGQYITDGVVGNFDGLKSCMEFGSTWVSPRTSAE 171
QY 192 NLIADYVAKSGFF--RVFVIGVNDLREGEYMTDNTPLQYNNWNGEPSPDPYGHEDCV 249
Db 172 NOALLKLVWSSGLSSKPKYIGVTDRETEGRFVNTGKQL-TFTNMGCPQPDYKGLQDCG 230
QY 250 EMLSSGRWNTTECHLTMFVCEFIKK 275
Db 231 VIEDSGLDGSGDIRPIMCEIDNK 256

RESULT 10
Q91907
ID Q91907 PRELIMINARY; PRT; 246 AA.
AC Q91907;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MANNOSE BINDING-LIKE LECTIN PRECURSOR (FRAGMENT).
GN MBL.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjold K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.";
RL Immunogenetics 51:955-964(2000).
DR EMBL; AF227739; AAF63470.1; -.
DR HSSP; P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS0041; C_TYPE_LECTIN_2; 1.
KW Signal; Lectin.
FT SIGNAL 1 13 POTENTIAL.
FT VARIANT 145 145 S -> F.
SQ SEQUENCE 246 AA; 25709 MW; AB692282D289D0D5 CRC64;

Query Match 20.7%; Score 307; DB 13; Length 246;
Best Local Similarity 31.7%; Pred. No. 2.6e-20;
Matches 83; Conservative 39; Mismatches 116; Indels 24; Gaps 8;

QY 14 ILLVFLLIQISGLIDISDPTAEVCATHTISPGKGDGEGKKGKVGKMGPK 73
Db 1 LLLVFLLIQISGLIDISDPTAEVCATHTISPGKGDGEGKKGKVGKMGPK 52
QY 74 GIKGELGDMGDRGNIGKTPTGKKGDKGKGLLIGIPGKAGTVCDCGRYKFKVGLDI 133
Db 53 GEKGESG-VSVQPGPKAGPAGTAGEKGERGSGGQSGPGS-----ESVLESLS 101
QY 134 STARLKTSMKFKVKNV--IAGIRETEEFYIVQEEKNYRESLTHCRIGMLAMPKDEA 191
Db 102 ETQQLAKIATEFEKVSCHFRKVGQYITDGVVGNFDGLKSCMEFGSTWVSPRTSAE 161

```

QY 192 NLLIADYVAKSGF--FRVFIGVNDLEREGQYMTNTPLQYNSWNEGSPDPYGHEDCV 249
Db 162 NQALIKLVSSGLSKKPYIGVTRKTEGQVDETKQL-TFTNWGPGQDDYKGLQDCG 220
QY 250 EMLSSGRWNTDECHLTMFVCE 271
Db 221 VIEDGLWDDGGCDIRPMCE 242

RESULT 11

Q28518
ID Q28518 PRELIMINARY; PRT; 236 AA.
AC Q28518;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE MANNOSE-BINDING PROTEIN C (FRAGMENT).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Medline-97031450; PubMed-8877375;
RT "Mogues T., Ota T., Fauber A.I., Sastry K.N.;
"Characterization of two mannosyl-binding protein cDNAs from rhesus
monkey (Macaca mulatta): structure and evolutionary implications.";
RL Glycobiology 6:543-550(1996).
DR EMBL; L43911; AAB48071.1;
DR HSSP; P11226; 1HUP.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
FT NON TER 1
SQ SEQUENCE 236 AA; 24911 MW; 449CBEB87C89732E CRC64;

Query Match 20.0%; Score 297.5; DB 6; Length 236;
Best Local Similarity 32.0%; Pred. No. 1.9e-19;
Matches 78; Conservative 33; Mismatches 110; Indels 23; Gaps 7;
QY 38 VCATHITSPGKDDGK-----GDPGEKKGKVGKGMKGIKGLGMDGRGNIGKGTG 93
Db 5 ICTNSTGGPQHADSPQVACNSPGINGFGKDGKDGKGEKGEFG-QGLRGLQGPCK 63
QY 94 IGKKDKGKGLGIPGEKAGTVDC-----GRYKFGVGLDISIARKTKSMKTKVNV 148
Db 64 LGPGNPGSSGPGKQKGDGPGSDCESSLAASERK---ALOTEMARKKWLTSGLG- 119
QY 149 IAGIRETEKFFYIVQEKYRESLTHCRIRGGMAMPKDEAANTLIADYVAKSGFRVF 208
Db 120 ----RQVGNKFLINGEMWTFDKVLCALCAQASVATPNAENAIOLNKE-----EAF 171
QY 209 IGVNLEREGQYMTNTPLQYNSWNEGSPDPYGHEDCVLSSGRWNTDECHLTMF 268
Db 172 LGTIDTETGEFVLTGNKL-TVTNWDDGPNAGSNEDCVLLKNGKWNIDPCSSSLA 230
QY 269 VCEF 272
Db 231 LCEF 234

RESULT 12

Q9T06
ID Q9T06 PRELIMINARY; PRT; 248 AA.
AC Q9T06;
DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PULMONARY SURFACTANT PROTEIN A (PULMONARY SURFACTANT-ASSOCIATED
DE PROTEIN A).
GN SPAS OR SP-A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20215262; PubMed=10749753;
RA Braams G.A., Yao L.-J., Inchley K., Brickenden A., Han V.K.M.,
RA Grolla A., Challis J.R.G., Possmayer F.;
RT "Ovine surfactant protein CNAs: use in studies on fetal lung growth
RT and maturation after prolonged hypoxemia.";
RL Am. J. Physiol. 278:L754-L764(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=20215263; PubMed=10749754;
RA Pietschmann S.M., Pison U.;
RT "CDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of
RT two different sequences for SP-B.";
RL Am. J. Physiol. 278:L765-L778(2000).
DR EMBL; AF211856; AAF18995.1;
DR EMBL; AF076633; AAF31148.1;
DR HSSP; P19999; 1YTT.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 1.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 248 AA; 26394 MW; D65E7293BBFF1FD9 CRC64;

Query Match 19.9%; Score 296; DB 6; Length 248;
Best Local Similarity 31.7%; Pred. No. 2.7e-19;
Matches 85; Conservative 34; Mismatches 113; Indels 36; Gaps 10;

QY 16 LVLFLLQISGLDIDSRPTAEVCATHTISPGKGDGKGDGKGVKGMKGI 75
Db 6 LFLMLNVAASGLECD---TREVCLG---SPGIPGPGHGLPGRDGRGKDPGPPG 59
QY 76 KCELDGM-GDRGNIGKTPGKKGKGLGIPGEKAGTVDCGRYKFGVGLDIS 134
Db 60 MGPPGPGPLPGRDGMGAPGLPGERGEK---GEPGERGPPG-----FPAYLD-- 104
QY 135 IARLTKSMKFKVKNVI-----AGIRETEKFFYIVQEKYRESLTHCRIRGMLA 184
Db 105 --EEQGLTHIRHOVLQSQGVLLILOGSMLEVGKVFSTNGOSLNFDAIKELCARAGHIA 163
QY 185 MPKDEAANTLIADYVAKSGFRVFIGNVNDLEREGQYMTNTPLQYNSWNEGSPDPY 244
Db 164 APRSPNEEAITSVKKNTY-AVLGLAEGPTAGFYVLDGAPV-NYTNWYGPGRG-RG 220
QY 245 HEDCVEMLSGRWNTDECHLTMFVCE 272
Db 221 KEKCVETDGMNDKNCLOYLAICEF 248

RESULT 13

Q9N0G1
ID Q9N0G1 PRELIMINARY; PRT; 248 AA.
AC Q9N0G1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PULMONARY SURFACTANT PROTEIN A.
OS Equus caballus (Horse).


```
Db 132 -KQVGNKFELTNGEINTFEKVKALCVKQASVATPRNAENGAIQNLKE----EAFIGI 186
QY 212 NDLEREGQYMFDTNPLQNYSNWNEGEPSPYGHEDCVEMISSGRWNTDECHLTWYFVCE 271
Db 187 TDEKTEGQFVDLTGNEL-TYTNWNEGEPNAGSDEDCVLLKNGQWNVPCSTSHLAYCE 245
QY 272 F 272
Db 246 F 246
```

Search completed: July 3, 2002, 12:40:21
Job time: 360 sec

Date is not good

CC human collectin protein and its encoding polynucleotide. The human
CC collectin exhibits antibacterial and antiviral activity and can be used
CC as an agent for the treatment of human bacterial and viral infections.
CC This sequence represents the novel human collectin.
XX
SQ Sequence 277 AA;

Query Match 100.0%; Score 1484; DB 20; Length 277;
Best Local Similarity 100.0%; Pred. No. 4.7e-142;
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNGFASILLRNQFILLVFLQLQSLGLDIDSRPTAEVCATHISPGKGGDGEKGPGE 60
|||||
Db 1 mngfasillrnqfllvflqlqslglididsrptaevcathtispkpgdgdgkpgge 60
|||||
QY 61 EGKGGKGYRMPGKGGKGLGDMGRGNIGKTGPIGKDKDGEKGLGIPGKAGVIVCD 120
|||||
Db 61 egkbgkgyrmgpgkglgldmrgnigktgpiqkdkdgekglgipgekagvtvcd 120
|||||
QY 121 GRYRKEFGVGLDLSIARLNTSMKFKVNVIAIGIRETEEFYIVQEEKNYRESLTHCRIRG 180
|||||
Db 121 gpyrkfvgldisarlntsmkfkvnviagireteeefyivqeeeknyreslthcrirg 180
|||||
QY 181 GMLAMPDEANLTIADIVAKSGRPFVIGVNDLEREGQYNTDPTLQYNSNNWNEGEPS 240
|||||
Db 181 gmlampdeanltiadyvaksgfrfvgvndlereregymftdntplqynsnwnegeps 240
|||||
QY 241 DPYGHEDCEVMSLSSGRWNTDTECHTMVYFCEFIKKKK 277
|||||
Db 241 dpyghedcevmssgrwntdtechtmyvfcfeikkkk 277
|||||

RESULT 2
AAY41698
ID AAY41698 standard; Protein; 277 AA.
XX
AC AAY41698;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PR0702 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN W09946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079669.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.

XX PN WO9946281-A2.

PF	08-MAR-1999;	99WO-USO5028.
XX		
PR	10-MAR-1998;	96US-0077450.
PR	11-MAR-1998;	98US-0077632.
PR	11-MAR-1998;	98US-0077631.
PR	11-MAR-1998;	98US-0077649.
PR	12-MAR-1998;	98US-0077791.
PR	13-MAR-1998;	98US-0076004.
PR	17-MAR-1998;	98US-0040220.
PR	20-MAR-1998;	98US-0078886.
PR	20-MAR-1998;	98US-0078910.
PR	20-MAR-1998;	98US-0078936.
PR	25-MAR-1998;	98US-0078939.
PR	26-MAR-1998;	98US-0079294.
PR	26-MAR-1998;	98US-0079656.
PR	27-MAR-1998;	98US-0079663.
PR	27-MAR-1998;	98US-0079664.
PR	27-MAR-1998;	98US-0079689.
PR	27-MAR-1998;	98US-0079728.
PR	30-MAR-1998;	98US-0079786.
PR	30-MAR-1998;	98US-0079920.
PR	31-MAR-1998;	98US-0079923.
PR	31-MAR-1998;	98US-0080105.
PR	31-MAR-1998;	98US-0080107.
PR	31-MAR-1998;	98US-0080165.
PR	01-APR-1998;	98US-0080194.
PR	01-APR-1998;	98US-0080327.
PR	01-APR-1998;	98US-0080328.
PR	01-APR-1998;	98US-0080333.
PR	08-APR-1998;	98US-0080334.
PR	08-APR-1998;	98US-0081049.
PR	08-APR-1998;	98US-0081070.
PR	08-APR-1998;	98US-0081071.
PR	09-APR-1998;	98US-0081195.
PR	09-APR-1998;	98US-0081203.
PR	09-APR-1998;	98US-0081229.
PR	15-APR-1998;	98US-0081817.
PR	15-APR-1998;	98US-0081838.
PR	15-APR-1998;	98US-0081952.
PR	15-APR-1998;	98US-0081955.
PR	21-APR-1998;	98US-0082568.
PR	21-APR-1998;	98US-0082569.
PR	22-APR-1998;	98US-0082700.
PR	22-APR-1998;	98US-0082704.
PR	23-APR-1998;	98US-0082804.
PR	23-APR-1998;	98US-0082767.
PR	27-APR-1998;	98US-0082796.
PR	28-APR-1998;	98US-0083336.
PR	29-APR-1998;	98US-0083322.
PR	29-APR-1998;	98US-0083392.
PR	29-APR-1998;	98US-0083495.
PR	29-APR-1998;	98US-0083496.
PR	29-APR-1998;	98US-0083499.
PR	29-APR-1998;	98US-0083500.
PR	29-APR-1998;	98US-0083545.
PR	29-APR-1998;	98US-0083594.
PR	29-APR-1998;	98US-0083598.
PR	30-APR-1998;	98US-0083599.
PR	05-MAY-1998;	98US-0083742.
PR	06-MAY-1998;	98US-0084366.
PR	06-MAY-1998;	98US-0084414.
PR	07-MAY-1998;	98US-0084431.
PR	07-MAY-1998;	98US-0084598.
PR	07-MAY-1998;	98US-0084600.
PR	07-MAY-1998;	98US-0084627.
PR	07-MAY-1998;	98US-0084637.
PR	07-MAY-1998;	98US-0084639.
PR	07-MAY-1998;	98US-0084640.
PR	13-MAY-1998;	98US-0084643.
PR	13-MAY-1998;	98US-0085323.
PR	13-MAY-1998;	98US-0085338.

QY	1	cagcaatgaatgcttggcatcctgtcttcgaagaaccaattatcctcgtactat	60
Db	17	cagcaatgaatgcttggcatcctgtcttcgaagaaccaattatcctcgtactat	76
QY	61	ttctttgcaaatcgaagctcgtgggtcctggatattgatagcgcgtcctaccgctgaagctc	120
Db	77	ttctttgcaaatcgaagctcgtgggtcctggatattgatagcgcgtcctaccgctgaagctc	136
QY	121	gtgccacacacaaattccaccagacccaaggagatgatggtgaaaaggagatccag	180
Db	137	gtgccacacacaaattccaccagacccaaggagatgatggtgaaaaggagatccag	196
QY	181	gagaagaggaaaagcatggccaaagtggacgcattgggccgaagaagaaataaaggagaac	240
Db	197	gagaagaggaaaagcatggccaaagtggacgcattgggccgaagaagaaataaaggagaac	256
QY	241	tggtgatattgggagatcctgggcataattggcgaagactgggccattgggaagaagggtg	300
Db	257	tggtgatattgggagatcagggaataattggcgaagactgggccattgggaagaagggtg	316
QY	301	acaaaggggaaaaaggttgcttggataactcgtggagaaaaggcaaacgagctactgct	360

BAKE
pan 0
N-PSDB; AA
WP: 2001-
03-2802

DI 8 10:13:15 2002

us-09-600-932-2.rag

SEQ 1D NO: 2

File COTM
Page 5
AC NO: AAY79510, Database: A-Genesys-
03-2802

XX PD 14-SEP-2000.
XX FF 06-JAN-2000; 2000WO-US00376.
XX PR 08-MAR-1999; 99WO-US05028.
XX PR 02-JUN-1999; 99WO-US12252.
XX PR 23-JUN-1999; 99US-0141037.
XX PR 07-JUL-1999; 99US-0143048.
XX PR 26-JUL-1999; 99US-0145698.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PA (GETH) GENENTECH INC.
XX FI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
XX FI Watanabe CK, Wood WI;
XX DR WP1: 2000-572070/53.
XX DR N-PSDB; AAC58385.
XX CC Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
XX FT treatment, diagnosis and prevention of cancer -
XX PS Claim 61; Fig 38; 286pp; English.
XX CC The present invention describes an isolated antibody that binds to
XX CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
XX CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO1005, PRO1009,
XX CC PRO1025, PRO1030, PRO1037, PRO1107, PRO1153, PRO1182, PRO1184,
XX CC PRO1187, PRO1281, PRO233, PRO339, PRO834, PRO1317, PRO1710, PRO2094,
XX CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
XX CC growth. The PRO polypeptides and nucleotides are useful in the
XX CC treatment, diagnosis and prevention of cancer. The antibodies and other
XX CC anti-tumour compounds may be used to treat various conditions, including
XX CC those characterised by overexpression and/or activation of the amplified
XX CC PRO genes. Exemplary conditions or disorders to be treated with such
XX CC antibodies and other compounds include benign or malignant tumours
XX CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
XX CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
XX CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
XX CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
XX CC glial, astrocytic, hypothalamic and other glandular, macrophagal,
XX CC epithelial, stromal and blastocoele disorders, and inflammatory,
XX CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
XX CC primers and hybridisation probes used in the isolation of the human PRO
XX CC sequences. AAC58367 to AAC58396 and AAC24057 to AAC24089 represent human
XX CC PRO polynucleotide and protein sequences given in the exemplification of
XX CC the present invention.
XX CC Sequence 271 AA;
XX SQ

Query Match 48.3%; Score 716.5; DB 21; Length 271;
Best Local Similarity 47.4%; Pred. No. 3e-64;
Matches 129; Conservative 60; Mismatches 76; Indels 7; Gaps 3;
QY 8 LRNFQFILLVFLIQSLQGLDIDSRPTA-EVCATHTISPGPKGDGKGPGECKHCK 66
DB 1 mrgulalvgvlislaflslpsghpqpagddacsqilnqglkgdagekd---kgatgr 57
QY 67 VGRMGPKFKGELGDMGDGNI---GKTGPTRKKGDKGKQLLGPGEKKGATGVCDCGR 123
DB 58 pgrvgtgkdgmdkgksgsvrghkigpigsqksgdsgdpppnpgeplpcesq 117
QY 124 YRKEVQLDISIARLTKSMFKVKNVAGIRETEKFTYIVQEKYRESLTHCRIRGML 183
DB 118 lraigemndqvqsltselkfknavagvreteskiylvkeekydaqlscqgggl 177
QY 184 AMPKDEAANTLADYVAKSFFRVFVNDLREGQYMTDTNPLQNSWNGESDSFY 243
DB 178 smpkdeanglmaaylaqaglarlvfigndlekegafvysdshspmrtnwrsgeppnny 237

QY 244 GHEDCVMLSSGRWNTDECHLTMYFVCEFIKK 275
DB 238 deedcvemvsggwndvachtmyfmcefdke 269
RESULT 6
ID AAY79510
XX ID AAY79510 standard; Protein; 271 AA.
XX AC AAY79510;
XX DT 01-AUG-2000 (first entry)
XX DE Human carbohydrate-associated protein CRBAP-6.
XX KW CRBAP-6; carbohydrate-associated protein 6; human;
XX KW autoimmune disorder; inflammation; gastrointestinal disorder;
XX KW infection; reproductive disorder; neurological disorder;
XX KW eye disorder; cell proliferation; cancer; diagnosis; therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20 /note= "signal peptide"
FT Protein 21..271 /note= "mature protein"
FT Modified-site 90 /note= "O-phosphorylated"
FT Modified-site 154 /note= "O-phosphorylated"
FT Domain 41..112 /note= "collagen-like domain"
FT Domain 247..256 /note= "C-type lectin domain"
FT Domain 44..77 /note= "C1q domain protein sequence"
FT Domain 150..265 /note= "C-type lectin family sequence"
XX WO2000018922-A2.
XX 06-APR-2000.
XX 29-SEP-1999; 99WO-US22685.
XX 01-OCT-1998; 98US-0164785.
XX 06-OCT-1998; 98US-0167179.
XX 13-NOV-1998; 98US-0191838.
XX 03-DEC-1998; 98US-0205656.
XX 03-DEC-1998; 98US-2223333.
XX (INCY-) INCYTE PHARM INC.
XX AU-Young J, Lal P, Bandman O, Reddy R, Baughn MR, Yue H;
XX Hillman JL;
XX WPI: 2000-317516/27.
XX N-PSDB; AA294946.
XX Novel carbohydrate-associated proteins used for the prevention and
XX treatment of autoimmune/inflammatory disorders of e.g. the
XX gastrointestinal and reproductive systems -
XX Claim 1; Page 88; 104pp; English.
XX The present sequence is that of a novel human carbohydrate-associated
XX protein, termed CRBAP-6, as deduced from cDNA (see AA294946), obtained
XX from adrenal tumour cDNA library ADRETUT06. CRBAP-6 has chemical
XX and structural similarity to bovine lung surfactant protein D (32%
XX identity). CRBAP-6 is expressed in liver, kidney, ovary, gut,
XX adrenal gland and secretory epithelium tissues. The invention

human collectin protein and its encoding polynucleotide. The human collectin exhibits antibacterial and antiviral activity and can be used as an agent for the treatment of human bacterial and viral infections. This sequence represents the novel human collectin.

Sequence	277 AA;
SQ	

Query Match 100.0%; Score 1484; DB 20; Length 277;
Best Local Similarity 100.0%; Pred. No. 4.7e-142;
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1	MNGFASLARNQFILLVFLQIQSLUGLIDISRPTAEVCATHITSPGPKGGDDGKGDGPE	60
	db	
1	mngfaslarnqfllvflfqsluglidisrptaevcathtispgpkggdgdgkdgpe	60
	db	
61	EGKHGKVGKMGPKGKXGELGMDGRNIGKTPGKGGKDGKGGLLGTPGKGGKAGTVCD	120
	db	
61	egkhgkvgrmpkgkxgelmgdgrnigktpgkggkdgkggllgtpgkggkagtvcd	120
	db	
121	CGRYKRFVQGLDISIARLTKSMKFKVKNVAGIRETEEKFYIVGEEKNYRESLTHCRIRG	180
	db	
121	cgryrkrfvgqldisiarltsmkfkvnvagiireteekfyivgEEKnyreslthcrrg	180
	db	
181	GLMLPKDEAANTLIADYVAKSGFRFVIGVNDLEREGQVMTDNTPLQNYSNWNGEFS	240
	db	
181	gmlmpkdeaanliadyvaksgfrfvigvndlerEGqvmtdntplqnySNwngEfs	240
	db	
241	DPYGHEDCVEMLSGGRWNTECHLTMYVCFIKKKK	277
	db	
241	dpynghedcvemlssgrwndtEchltmyvCFIkkkk	277

RESULT 2

AA41698

ID AAY41698 standard; Protein; 277 AA.

AA41698:

[illegible]

07-DEC-1999 (first entry)

Human PRO702 protein sequence.

General

Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation; probe; blood coagulation disorder; cancer; cellular adhesion disorder; secreted protein; transmembrane protein.

F40000

Homo sapiens.

NY
DN
LX-16634660M

W09940281-A2..

16-SEP-1999.

10

08-MAR-1999;

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

10-MAR-1998;
11-MAR-1998

II-MAR-1998;
11-MAR-1998.

11-MAR-1998;
17-MAR-1999.

12-MAR-1998:

13-MAR-1998;

17-MAR-1998;

R 20-MAR-1998;

R 20-MAR-1998;
P 20-MAR-1998;
P 20-MAR-1998;

R 20-MAR-1998;
R 20-MAR-1998;
R 20-MAR-1998;

R 20-MAR-1998;
R 25-MAR-1998;

K R
23-MAR-1998;
26-MAR-1998:

20 MAR 1998,
27-MAR-1998:

R 27-MAR-1998;

27-MAR-1998;

27-MAR-1998;

27-MAR-1998;

3

19

20

1

PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 XX WPI; 1999-551358/46.
 DR N-PSDB; AA233973.
 XX
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders.
 XX
 PS Claim 12; Fig 37; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AA233891 to
 CC AA234338, and AA41685 to AA41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 277 AA;
 Query Match 99.2%; Score 1472; DB 20; Length 277;
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNGFASLLRRNQIFLLVFLQLQISGLDIDSRPTAEVCATHITSPGKGDGEGDPE 60
 Db 1 mgfaslrrnqifllvflqlqisglididsrptaevcathtispqkdgdekdpge 60
 QY 61 EGKHGKVGKMGPKIGKELGMDGRNIGTKTGPIGKKGKGLLIGIPGKAGTVC 120
 Db 61 egkhgkvgrmgpkigkglmdgdnigtktpigkkgkgligipgkagtvcd 120
 QY 121 CGRYKFKVQGLDISIARLTKSMKFKVKNVIAGIRETEEFYIVQEKNYRESLTHCRIRG 180
 Db 121 cgryrkfvqglidisarltsmkfkvnviagireteekfyivqeknyreslthcrirg 180
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFVIGVNDLREGQYMTNTPLQYNSWNEGEPS 240
 Db 181 gmlampkdeaanliadyvaksgffrfvignvdleregymntntplqynswnegeps 240
 QY 241 DPYGHEDCEVMLSSGRWNDECHLTMYFVCEFIKKKK 277
 Db 241 dpyghedcvmllssgrwndtechltmyfvcfeikkkk 277

RESULT 3
 AAB44554
 ID AAB44254 standard; Protein: 277 AA.
 XX
 AC AAB44254;
 XX
 XX 08-FEB-2001 (first entry)
 XX
 DE Human PR0702 (UNQ366) protein sequence SEQ ID NO:97.
 XX
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
 KW expressed sequence tag; detection; cancer.
 XX
 OS Homo sapiens.
 XX
 PN W0200053756-A2.
 PD 14-SEP-2000;
 XX
 PF 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.

AR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 19-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 05-JAN-2000; 99WO-US31274.
 PR 06-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi C, Gurney AL, Hillan KJ;
 PI Kljavin LJ, Luo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 DR WPI; 2000-611443/58.
 DR N-PSDB; AAC78480.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities.
 XX
 PS Claim 12; Fig 37; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provided specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78897 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 XX
 SQ Sequence 277 AA;
 Query Match 99.2%; Score 1472; DB 21; Length 277;
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNGFASLLRRNQIFLLVFLQLQISGLDIDSRPTAEVCATHITSPGKGDGEGDPE 60
 Db 1 mgfaslrrnqifllvflqlqisglididsrptaevcathtispqkdgdekdpge 60
 QY 61 EGKHGKVGKMGPKIGKELGMDGRNIGTKTGPIGKKGKGLLIGIPGKAGTVC 120
 Db 61 egkhgkvgrmgpkigkglmdgdnigtktpigkkgkgligipgkagtvcd 120
 QY 121 CGRYKFKVQGLDISIARLTKSMKFKVKNVIAGIRETEEFYIVQEKNYRESLTHCRIRG 180
 Db 121 cgryrkfvqglidisarltsmkfkvnviagireteekfyivqeknyreslthcrirg 180
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFVIGVNDLREGQYMTNTPLQYNSWNEGEPS 240
 Db 181 gmlampkdeaanliadyvaksgffrfvignvdleregymntntplqynswnegeps 240
 QY 241 DPYGHEDCEVMLSSGRWNDECHLTMYFVCEFIKKKK 277
 Db 241 dpyghedcvmllssgrwndtechltmyfvcfeikkkk 277

PD 16-SEP-1999.

QY	1	cagcaatgaatggtcttgcattcttgcctcgaagaaaccaatttactctctcgtgactat	60
Db	17	cagcaatgaatggtcttgcattcttgcctcgaagaaaccaatttactctcgtgactat	76
QY	61	ttctttgcaattccagagtctgggtctgattgatagccctactaccctgaagtct	120
Db	77	ttctttgcaattccagagtctgggtctgattgatagccctactaccctgaagtct	136
QY	121	gtgccacacacaatttcacacaggaccacaaaggagatgatggtgaaaaggagatccag	180
Db	137	gtgccacacacaatttcacacaggaccacaaaggagatgatggtgaaaaggagatccag	196
QY	181	gagaagagggaaagcatggtcaaaagtgggacgcatggggccgaaaggaaattaaaggagaac	240
Db	197	gagaagagggaaagcatggtcaaaagtgggacgcatggggccgaaaggaaattaaaggagaac	256
QY	241	tgggtatattgggagatcggggccaattattggcaagactgggccattgggaagaagggtg	300
Db	257	tgggtatattgggagatcagggccaattattggcaagactgggccattgggaagaagggtg	316
QY	301	acaaaggggaaaaaggttcttggaataacctcgagaaaaaggcaagcaggtactgctt	360

